# LARGE PLANT DATA IN THE LRD: SELECTION OF A SAMPLE FOR ESTIMATION

**by**

**Phoebus J. Dhrymes**
**Columbia University**

**and**

**Linda Moeller**
**U.S. Bureau of Labor Statistics**

**CES 99-6 March, 1996 (revised April, 2002)**

The research upon which this paper is based was carried out, in part, during the authors' participation in the ASA/NSF/Census Research Program; Dhrymes served as a Senior Research Fellow and Moeller served as his Research Associate. The opinions, findings and conclusions or recommendations expressed in this paper are those of the authors and should not be attributed to the American Statistical Association, Columbia University, the National Science Foundation, the U. S. Census Bureau, the Census Bureau's Center for Economic Studies, the U. S. Department of Commerce, the U.S. Bureau of Labor Statistics, the Bureau of Labor Statistics' Office of Productivity and Technology, Division of Productivity Research, or the U. S. Department of Labor. The paper has been screened to ensure that it does not disclose confidential information.

## Introduction

This paper describes preliminary work with the LRD during our tenure at the Census Bureau as participants in the ASA/NSF/Census Research Program.[1] The objective of the work described here were two-fold. First, we wanted to examine the suitableness of these data for the calculation of plant-level productivity indexes, following procedures typically implemented with time series data. Second, we wanted to select a small number of 2-digit industry groups that would be well suited to the estimation of production functions and systems of factor share equations and factor demand forecasting equations with system-wide techniques. This description of our initial work may be useful to other researchers who are interested in the LRD for the analysis of productivity growth and/or the estimation of systems of factor equations, because the specific results reported in this memo suggest that the data are of good quality, or

---

[1] Dhrymes served as an ASA/NSF/Census Research Fellow on a half-time basis during the 1988-89 academic year, and on a more limited basis during the 1989-90 academic year. Moeller served full-time as his ASA/NSF/Census Research Associate. We are indebted Harold Watts for recommending that we work with the LRD, and to to the Census Bureau's Statistical Research Division (SRD), which housed the ASA/NSF/Census Research Program, for their support and hospitality. We would particularly like to acknowledge the extraordinary computer support provided by Neal Bross and Chris Dyke of the SRD.

The empirical work described here was completed during the first year of our tenure at the Census Bureau. Unfortunately, time constraints prevented us from documenting our preliminary empirical work during the period of our formal participation in the ASA/NSF/Census Research Program. We are indebted to the Division of Productivity Research, in the Bureau of Labor Statistics' Office of Productivity and Technology, for permitting Moeller to spend part of her research time drafting this report and for storing computer output that was generated during the course of the project.

because the nature of the tasks undertaken provides insight into issues that arise in the analysis of longitudinal establishment data.

Two sets of results reported in this paper are especially noteworthy for the purposes of conventional index number calculations, of the sort that the Bureau of Labor Statistics employs to measure growth in multifactor productivity. First, as discussed in Sections I and IV below, year-to-year changes in the population of plants within a two-digit industry are often quite small, averaging 1-3% of all plant-year observations among the industries that we studied. Plants with 250 or more employees are certainty cases in the Annual Survey of Manufactures and the Census of Manufactures, which collect the original microdata. The measurement error associated with these observations consists entirely of non-sampling error.[2] Thus these large

---

[2] The formula for the variance of a population mean for the variable y, when estimated with data from a stratified sample, is as follows:

$V'y_{st} = (1/N) \sum_L N_h' \; N_h - n_h \; ( S_h^2/n_h ),$ where
- N is the size of the population;
- $N_h$ is the size of the population in stratum h;
- $n_h$ is the size of the sample in stratum h;
- $S_h^2$ is the population variance in stratum h;
- $N_h - n_h = 0$ for certainty cases, and
- $h = 1,\ldots,L$.

This formula is from W. G. Corcoran (1977), <u>Sampling Techniques</u>, Third Edition, New York: John Wiley, Chapters 5 and 5A.

plant data are the most reliable component of growth rate estimates based on the published industry time series.[3]

A second result that is particularly important for index number calculations, discussed in Section V, is the feasibility of constructing plant-specific capital stock estimates according to the perpetual inventory approach, following a simplified version of the the approach followed by the BLS. Data on building investment proved too "lumpy" to permit the construction of separate capital stock estimates for plant and equipment.[4] However distributions of the plant-stock estimates for total capital stock were quite skewed, and consistent with the distributions of employment, value added, and shipments.

For the purposes of estimating the parameters of systems of factor share equations, which might in turn be used to construct index number weights, the fact that imputation rates are low among observations on plants with 250 or more employees, as discussed in section IC below, is particularly noteworthy. Measurement error can affect all estimated parameters when system-

[3] Repeated observations on individual units of analysis improve the efficiency of estimated changes for the population as a whole because they decrease the share of the estimated change that is attributable to sampling variation. See Corcoran (1977), cited above, especially p. 353.

[4] In future work it might prove worthwhile to to estimate a smoothed plant-specific building stock series, perhaps by applying discrete categorical variable estimation procedures such as Tobit or probit to merged plant- and company-level data, recognizing that "bricks-and-mortar" investment investment decisions often are not taken at the plant level. Unfortunately such a project was well beyond the scope of our initial work.

wide techniques are employed. However true production function vaues are likely to vary systematically across sample strata, and when this is so the estimation of stratum-specific parameter estimates may be indicated. Thus the use of system-wide estimation techniques with large plant data, in combination with the application of simpler econometric techniques to samples drawn from other strata in the population, might yield particularly efficient estimates.[5]

Given the two goals listed above, and having undertaken the preliminary data analysis described below, we decided to work with LRD microdata from manufacturing plants with 250 or more employees whose primary product codes fell within major industry groups 35 (machinery except electrical), 36 (electrical machinery) or 38 (fine instruments) over the time period 1972-1986. Plants with 250 or more employees accounted for 80% of the total value of annual shipments in these industries over the period examined.

---

[5] Subsequent results from the estimation of systems of factor share equations derived from production and cost functions with these data, for the Cobb-Douglas, Constant Elasticity of Substitution and Translog functional forms, are presented in Phoebus J. Dhrymes (1990), "The Structure of Production Technology: Evidence from the LED Sample I," Bureau of the Census 1990 Annual Research Conference Proceedings, pp. 197-293. That paper reveals some inconsistencies between parameter estimates derived from cost functions and those derived from production functions. The results reported here suggest that these inconsistencies probably are not attributable to errors in the microdata. Additional substantive results based on these data are reported in Linda Moeller (1989), "On the Estimation of Systems of Long-Run Factor Demands," presented at a seminar of the Center for Economic Studies; Dhrymes (1991), "The Structure of Production Technology: Productivity and Aggregation Effects," Center for Economic Studies Discussion Paper CES 91-5, Eric J. Bartlesman and Phoebus J. Dhrymes (1992), "Productivity Dynamics: U. S. Manufacturing Plants, 1972-1986," Center for Economic Studies Discussion Paper CES 92-1; Moeller (1993), "Large Machinery Plants' Demands for Factors of Production," presented at the Annual Conference of the Western Economic Society International in Reno, Nevada; and Moeller (1995), Systems of Factor Demand Equations Derived from a Model of Monopolistic Competition: Results from Time Series Cross Section Data, Ann Arbor: University Microfilms.

The balance of the paper is organized chronologically, to reflect the decision-making process followed. Information reported by researchers with prior experience with the LRD, which led to the decision to work exclusively with data from large establishments, is summarized in Section I. Results from a first set of extracts that contained a small number of variables from all large manufacturing establishments are reported in Sections II-IV. Section II focuses on the compositional stability of 2- and 3-digit industry groups over the period 1972-1986: graphical evidence on the stability of large plants' shares in total shipments by 2- and 3-digit industry is described, as is graphical evidence on the stability of sub-industry shares among large plants. Detailed published results on imputation rates by 4-digit industry group are discussed in Section III, for selected 2-digit industries. Information on the frequency with which plants moved out of selected 2-digit industry groups, and on the duration of these spells within a given industry and size category, is presented in Section IV. Counts of the number of observations available annually within selected 2-digit industries and 3-digit sub-industries are also reported in Section IV.

Additional start-up tasks are described in Sections V and VI. Different industry price indexes that might be used to convert nominal values to "constant dollars," the construction of establishment-specific investment deflators, and the calculation of value added are reviewed briefly in Section V. Estimation of the real value of establishment-specific stocks of plant and equipment according to the "perpetual inventory technique" is described in Section VI. Distributions of these capital stock estimates are compared with distributions of total employment, the total value of shipments, and value added. The paper is summarized in Section VII.

## I. Prior Results: Relative Importance, Continuity, and Imputation

The microdata examined in this study were originally collected for the Annual Survey of Manufactures (ASM) and the Quinquennial Census of Manufactures (CM).[6] Annual microdata from the period 1972-1981 and Census microdata from the years 1963, 1967, 1972, 1977 and 1982 were assembled into the Longitudinal Establishment Data file by Nancy and Richard Ruggles of Yale University, in a joint project with the Census Bureau. This file has been extended to incorporate microdata from subsequent surveys and censuses, and renamed the Longitudinal Research Data file (LRD) in order to emphasize the fact that the data set has been assembled primarily for economic research.[7]

---

6 The materials discussed in Section 1 were provided to Moeller by James Monahan, formerly of the Census Bureau's Center for Economic Studies (CES) , in response to her initial inquiries about the LRD. Mr. Mohanan was an invaluable source of information about the LRD, and we are grateful to him for his advice. We are indebted to CES Chief Dr. Bob McGuckin, Assistant Chief Bob Bechtold, and the CES for their assistance and encouragement. Moeller also thanks the Office of Tax Policy of the New York City Department of Finance for supporting some of this initial work.

7 A concise introductory description of the LRD is provided in Robert H. McGuckin and George A. Pascoe, Jr. (1988), "The Longitudinal Research Database: Status and Research Possibilities," Survey of Current Business, November, pp. 30-37.

During the first month on the project, we decided to focus on the analysis of large plants, or "certainty cases" in the Annual Survey of Manufactures (ASM).[8] This decision was based on information on the relative importance of shipments from large plants in total industry shipments, on the continuity of coverage over adjacent survey years for different employment size categories, on response rates for different employment size categories, and on the realization that parameter estimates obtained with these observations would not be subject to sampling error.[9]

### A. Relative importance of large plants' shipments.

---

[8] The ASM is an optimally stratified sample of establishments. In optimally stratified samples certainty cases are included in the sample surveyed with a probability of one. The motivation for surveying large plants with certainty is two-fold. First, the resulting decrease in the sampling variability of estimates for population totals is judged to warrant the cost of surveying the entire upper tail of a highly skewed distribution. And second, the sub-population in the upper tail is believed to be so heterogeneous that a random sample would not be representative. To illustrate the second concern a statistician in the IRS has explained, "We want to be sure we include General Motors in the survey, and we want to be sure we don't count them twice." A lucid discussion of optimally stratified sample design is provided in Corcoran (1977), cited above.

[9] The term "sampling error" refers to errors of measurement due to the fact that one examines a random sample from the population in question, rather than from the population as a whole. Sampling error is a separate concept from that of selectivity bias: a sample whose observations are drawn randomly from an entire population is not subject to selectivity bias, but it is still subject to sampling error.

There is also a literature on nonsampling error as it arises in the analysis of longitudinal data. See Mark Roberts and James Monahan (1986), "The Effects of Nonsampling Errors on the Development and Use of the Longitudinal Establishment Data (LED) File," and Graham Kalton, David McMillen, and Daniel Kasprzyk (1986), "Nonsampling Error Issues in the Survey of Income and Program Participation (SIPP)," and the references provided in these papers. Both papers are published in the Bureau of the Census Second Annual Research Conference Proceedings, March 1986.

Ruggles and Ruggles report that there were a total of approximately 360,000 domestic manufacturing establishments in 1977.[10] Among all manufacturing establishments the roughly 43,000 certainty cases in the ASM, or 12 percent of the total number of plants in 1977, accounted for 80 percent of the aggregate value of manufacturing shipments.[11] This extreme quantitative importance of a relatively small number of large establishments is characteristic of the manufacturing industry.

## B. Continuity of coverage over survey years

LRD data from large plants are relatively well suited for work that involves dynamic models because data from large plants are readily linked over time.[12] This greater continuity of coverage for plants with 250 or more employees is largely due to the fact that they are certainty cases in the ASM. Ruggles and Ruggles report that, for the period 1974-81, the percentages

---

[10] An establishment is a plant or other physical location where productive activity takes place. Establishments are to be contrasted with enterprises, or firms. A single enterprise may own one or more establishments.

[11] Nancy Ruggles and Richard Ruggles (1984), "The Analysis of Longitudinal Establishment Data," unpublished mimeo cited with the permission of Richard Ruggles.

[12] For the purposes of this project, the term "large plants" identifies plants with 250 or more employees. Hsiao describes maximum likelihood procedures that may be followed to estimate systems of equations with data in which observations on some units of observation are missing for some periods spanned by the data set. These, or similar, procedures would be required to estimate dynamic models for the industry as a whole. See Cheng Hsiao (1986), Analysis of Panel Data, Cambridge: Cambridge University Press.

of the total value of shipments attributable to plants that could be linked over time, relative to the total value of shipments attributable to the entire employee-size class, are as shown in Table 1.[13] These percentages increase steadily from less than 19%, in the case of plants with fewer than 100 employees, to over 90% in the case of establishments with 500 or more workers.

### C. Imputation Rates

It is well known that measurement errors may be propagated throughout an entire set of parameter estimates with the implementation of system-wide estimation procedures such as 3SLS and full-information maximum likelihood. For this reason, the fact that large plants are generally found to report more completely and more consistently than smaller survey respondents made large plants an especially attractive sub-population for our purposes.

To illustrate the lower imputation rate observed in data from large establishments, Ruggles and Ruggles report the ratios shown in Table 2. In these ratios the numerator is the total estimated value of shipments with imputed values omitted, and the denominator is the total with imputed

---

[13] Ruggles and Ruggles (1984), p. 14.

values included. [14] On average, the actual reported value of shipments (as opposed to the imputed value) accounts for more than 90% of the total estimated value of shipments from large plants. More detailed evidence that the percentage of surveyed establishments who complete survey questionnaires is usually higher among large establishments than it is among other employment-size categories is provided by Monahan and Roberts, and summarized in Table 3.[15] This table shows that non-imputed values account for more that 90% of the estimated value in the case of large plants, for almost all variables examined.

Having decided to specialize our analysis to a sub-population of large plants on the basis of work by prior researchers, we requested that all microdata from plants with 250 or more employees be extracted from the LRD and copied onto computer tapes for our use.[16] The next steps in our industry selection procedure required that we extract plants' permanent identification numbers (PPN), primary product code (IND), and total value of shipments (TVS)

---

[14] Ruggles and Ruggles (1984), p. 13. Imputed and missing values appear more frequently for variables other than the total value of shipments, and especially for variables not routinely reported in published ASM reports. See Table 3, below.

[15] Mark Roberts and James Monahan, "The Effects of Nonsampling Errors on the Development and Use of the Longitudinal Establishment Data (LED) File, photocopy. An abbreviated version of this paper is published in the 1986 Bureau of the Census Second Annual Research Conference Proceedings, pp. 131-146.

[16] This approach was adopted largely for simplicity, and to avoid unwarranted imposition on Center for Economic Studies (CES) staff members that might have resulted from repeated requests for smaller extracts. During the second month of the project CES staff member Jim Monahan kindly generated 15 extracts (one each year for the period 1972-1986) and copied them onto 8 tapes for our use. These extracts were the source files for all of our subsequent work on this project.

from the 199,873 annual observations on general economic variables (Record Type 11) from the 8 tapes that were created in response to our request, and store them on the virtual disks assigned to our project.[17] Over the period 1972-1986, observations on manufacturing establishments with 250 or more employees were distributed as shown in Table 4.

---

[17] We also read in the following variables from the detailed product and materials records from the records of 13,067 manufacturing plants with 250 or more employees in the 1982 Census-year extract, in order to assess the viability of detailed analysis using the product and materials data. From each detailed product record (Record Type 30) we extracted a product code (PI), and information on the quantity produced (PQP), quantity of shipments (PQS), value of shipments (PV), quantity of interplant transfers (PQIT), value of interplant transfers (PVIT), quantity produced and consumed (PQPC), and the total value of shipments (TVS). Inspection of a frequency distribution of product codes from the 66,687 detailed product records in the 1982 extract revealed 6,699 different 7-digit product codes. For each 6-digit materials record (Record Type 20), we extracted a materials code (M), and information on the quantity produced and consumed (MQPC), the quantity delivered and consumed (QPDC), and the cost of the materials (MC). A frequency distribution from the 89,470 detailed materials records included in the 1982 extract revealed 1,243 different materials codes.

CES staff members indicated that the first 6 digits of PI and the entire 6 digits of M are associated with the same sets of product characteristics in most cases, but that the correspondence between the two sets of definitions is not always exact. In light of the fact that the construction of establishment-level product and materials price indexes was not a primary objective of our project, we decided that a thorough analysis of the detailed product and materials data from 1972, 1977 and 1982 would absorb an undue amount of time. Subsequently we focused entirely on the general economic variables, which are available on an annual basis.

## II. Stability of Large Plants' Shares of Industry Shipments

We wanted to estimate production functions and factor demand equations from a population whose production possibility set appears to be relatively stable. Therefore our first extracts were used to plot the shares of total shipments attributed to large plants in each year within 2- and 3- digit groups, the shares of industry shipments attributed to large plants in each 4-digit group within the output attributed to large plants in each 3-digit group, and the shares of large plants' shipments attributed to the 3-digit groups within each 2-digit group. These charts provided us with preliminary evidence on the stability of the 4- and 3-digit composition of each 2-digit group.

If large plants within an industry do not constitute a separable subset of the total population of establishments, a model of the firm's choices regarding plant size and the composition of production should be incorporated in the systems of factor demand equations estimated in order to test for evidence of selection bias.[18] It is not obvious that these decisions can be modeled with establishment-level data alone unless the production function of the multi-establishment firm

---

[18] The structure of these tests is described in Phoebus J. Dhrymes (1986), "Limited Dependent Variables," in Handbook of Econometrics, Vol. III, Griliches and Intrilligator, Eds. , Amsterdam: North-Holland.

is separable in a partition by industry and plant size, because plant size and product composition decisions are typically made at the level of the firm rather than at the level of the establishment.[19] Therefore these charts were developed to gauge, albeit in a preliminary and informal manner, the appropriateness of an econometric analysis of the behavior of large plants separate from that of their parent firms and, implicitly, from that of other establishments owned by firms that operate in other industries or size classes.[20]

The first set of charts examines the stability of the total value of shipments attributed to large plants within a given 2- or 3-digit industry group, relative to the total value of shipments

---

[19] Explicit modeling of industry and establishment size decisions would permit the estimation of systems of equations that include industry switching equations. However, the relatively complicated nonlinear estimation procedures associated with switching equations were judged to be unduly complicated for initial work with a relatively new and unknown data set. This preference for a more straightforward initial approach to the data was reinforced by the fact that the firm-level microvariables one would want to employ in the specification of plant size and product switching equations were not then available for work with the LRD. The interpretation and expected signs of estimated factor demand elasticities for the multi-product, multi-plant firm are discussed by Ralph W. Pfouts (1961), "The Theory of Cost and Production in the Multi-Product Firm," Econometrica, 29, pp. 650-668, Pfouts (1964), "Some Cost and Profit Relationships in the Multi-Product Firm," Metroeconomica, 16, pp. 51-66, and Phoebus J. Dhrymes (1964), "The Monopolistic Multiproduct Firm Under Uncertainty," International Economic Review, 5, pp. 235-257. Estimation problems that may arise in the presence of individual firm effects are discussed in Moeller (1995), cited above.

[20] In the LRD, the permanent plant identification number (PPN) associated with each establishment can be used to identify establishments owned by multi-establishment firms. Before each economic census large companies are surveyed to obtain a list of the establishments that they own. Roberts and Monahan report, "Considerable effort is expended to verify the composition of large multi-unit companies since these companies account for over 80 percent of economic activity in terms of output, although they own only 10 percent of the number of establishments." See Mark Roberts and James Monahan (1986), "The Effects of Nonsampling Errors on the Development and Use of the Longitudinal Establishment Data File," Bureau of the Census Second Annual Research Conference Proceedings, Bureau of the Census, U. S. Department of Commerce, pp. 133.

attributed to all plants within that industry.[21] The second set of charts reviews the stability of the shares of shipments from large plants within each 2-digit group that are attributable to the large plants in each 3-digit subset, and the stability of the shares of large plants' shipments attributed to each 4-digit subset of large plants within all 3-digit groups. These charts provide graphical information about the stability of the composition of shipments within each subset.

### A. Shares of total industry shipments attributable to large plants.

The shares of total industry shipments attributable to large plants were calculated by summing the total value of shipments reported by the establishments in these initial extracts on an annual basis, by 3- and 4-digit category. Industry totals for large establishments were divided by the total value of shipments for the industry as a whole in the appropriate year, and the resulting shipment shares were plotted.[22] These plots provide preliminary evidence on the compositional

---

[21] The share of the establishment's total shipments attributable to its primary product is reported annually, in the primary product specialization ratio (PPSR). Detailed data on the composition of production are collected at 5-year intervals, in the quinquennial Census of Manufactures, but are not available in non-Census years. An informative investigation into the composition of products produced and materials consumed by establishments in the LRD is reported in Frank M. Gollop and James L. Monahan (1991), "A Generalized Index of Diversification: Trends in U.S. Manufacturing," Review of Economics and Statistics, 73, pp. 318-330. These indexes show significantly less product diversity at the establishment level than at the level of the firm, declining levels of diversity over time, and a negative relationship between product diversity and employment at the establishment level.

[22] We used the NBER Productivity Dataset maintained by Wayne Gray, mentioned above, as our source of data on the total value of shipments attributed to all establishments in each 3- and 4-digit group. Don

stability of the set of plants in each industry group within partitions defined by the total number of employees.

More specifically, let $n_{I,3,t}$ and $n_{I,4,t}$ represent the number of large plants within a 3- or 4-digit industry indexed $I$ at time t, let $S_{i,t}$ represent the value of shipments reported by a specific plant, and let $S_{I,3,t}$ or $S_{I,4,t}$ be the corresponding value of shipments attributed to the entire 3- or 4-digit industry. Then large plants' shares of total shipments were calculated as follows:

$$s_{I,j,t} = \frac{\sum_{i=1}^{n_{I,j}} S_{i,t}}{S_{I,j,t}}, \qquad j = 3,4; t = 72,\ldots,84.$$

The vertical axes of these plots range between 0 and 1, and the horizontal axes range between 72 and 84.[23] The 3-digit annual shares were plotted by 2-digit group and the 4-digit shares were plotted by 3-digit group. These plots of large plants' shipment shares are discussed in subsection II.A.1. The sub-group shares of all output produced by large plants in a given industry group were plotted similarly. Results from this second set of share plots are described in subsection II.A.2.

---

Siegel, who was just completing his tenure as an ASA/NSF/Census Research Associate with Frank Lichtenberg as our project began, kindly gave us a copy of this file.

[23] Although the microdata we obtained from CES span the period 1972-1986, the most recent year included in the file of time series aggregates given to us by Don Siegel was 1984. Consequently the first set of plots reported below, which are ratios of shipments from large plants in an industry to total industry shipments, were restricted to the period 1972-1984.

## 1. Large plants' shares of 3-digit industry shipments, by 2-digit groups

At the 2-digit level, plots of large plants' shares in total sub-industry shipments show that the relative rankings of the shares of output attributed to large establishments were fairly stable over time, although they were not entirely constant. For example, Figure 1 is a plot of large plants' shares among the 3-digit groups in industry 20. In this case the largest share of 3-digit shipments attributable to plants with 250 or more employees was that of industry 206: in industry 206 large plants were responsible for 65%-70% of total industry shipments over the period 1972-1984. Large plants were responsible for 55%-65% of the total value of shipments in industries 203 and 201, while the shipments shares of large plants in industry 205 were almost always slightly below those of industries 201 and 203. Large plants' shipment shares ranged between 45% and 50% for industry 208, and were always below those of the 3-digit groups already mentioned. Large plants' annual shares in industry 209 ranged between 35% and 45%, and were consistently smaller than the corresponding shares for industry 208. The shares of shipments attributed to large plants in industry 204 consistently ranged between 25% and 35%. For industries 202 and 207 large plants' shares ranged between 15% and 25%, with shares for industry 202 slightly below those of industry 207.

Within 2-digit groups, the largest share of shipments attributable to large plants is greater than 80% in all but industries 23, 24, 25 and 31. The 3-digit group in which large establishments account for the smallest share of total industry output have large plant shares that are almost always higher than 20% for all but industries 23, 24, 26, 29, 31, 32, and 34.[24] Distinctly larger large-plant shares are exhibited by industries 35 (where all 3-digit large plant shares usually range between 40% and 95%), 36 (range 60%-90%), 37 (range 70%-95%), and 38 (range 50%-90%).

[24] This statement does not always include cases in which the last digit is 9, which generally corresponds to the category "not elsewhere classified."

## 2. Large plants' shares of 4-digit industry shipments by 3-digit groups

At the 3-digit level, plots of large plants' sub-industry (4-digit) shares in total industry shipments show that the relative rankings of the shares of output attributable to large establishments are less stable over time than is the case at the 2-digit level. In some cases the sub-industry shares of two distinct sub-groups exhibit complementary fluctuations, as is the case for the 4-digit industries within the 3-digit group 234, shown in Figure 2. Although this pattern is unavoidable in the case illustrated in Figure 2 because the 3-digit group contains only two sub-groups, similar patterns are also observed between pairs of 4-digit groups within 3-digit industries when the total number of sub-groups is larger than 2. For example, 6 4-digit industries are included within the 3-digit group 238. But the shipments shares of the two sub-industries with the largest shares of total shipments within industry 238, industries 2381 and 2385, also exhibit complementary fluctuations, as illustrated in Figure 3. This pattern suggests that these two sub-industries may be ones in which the composition of output fluctuates systematically because the primary products of some multi-establishment plants fluctuate between 2381 and 2385.[25]

[25] In some cases individual multi-product plants produce roughly equal shares of output that is coded within different SIC categories, with the result that the primary product code assigned to the plant vacillates from year to year. In such cases Industry Division analysts may apply “resistance” to the assignment of primary product codes, in order to avoid spurious fluctuations in the industry time series. However, when the analyst believes that these fluctuations are substantive, they allow primary product codes to fluctuate. Therefore true fluctuations in product shares may be slightly understated in these charts.

## B. Shares of large plants' shipments attributable to industry sub-groups.

A second set of plots, discussed in this section, was obtained by summing the total value of shipments across all large plants within subsets of a given industry group, and across all large plants in that group. Sub-group shares were calculated as the ratios of the sub-groups sums to group totals. That is, using the notation defined in IIA., sub-group shares of large plants' total shipments were calculated as follows:

$$shr_{I,j+1,t} = \frac{\sum_{i=1}^{n_{I,j+1}} S_{i,t}}{\sum_{i=1}^{n_{I,j}} S_{i,t}}, \qquad j = 2,3; t = 72,\ldots,86.$$

These plots provide preliminary evidence regarding the stability of the composition of output within each industry sub-group in partitions defined by 2- and 3-digit SIC codes.[26] As was the case for the partition defined by number of employees, discussed in section A, sub-industry shares show a greater tendency to fluctuate at lower levels of aggregation. That is, 4-digit shares associated with 3-digit totals show a greater tendency to fluctuate in ranking than do 3-digit shares associated with 2-digit totals.

[26] A helpful discussion of the criteria according to which SIC code categories are defined is provided in James W. McKie (1985), "Market Definition and the SIC Approach," in Antitrust and Regulation: Essays in Memory of John H. McGowan, Franklin M. Fisher, Ed., Cambridge: MIT Press, pp. 85-100. We are indebted to James Monahan for this reference.

### 1. Three-digit shares of shipments from large plants, by 2-digit group

Figure 4 provides an example of results obtained by plotting the shares of total shipments from large plants in each 3-digit sub-group within 2-digit industries over the period 1972-1986. [27] The relative rankings of sub-groups shares for all large plants in industry 23 were remarkably stable over this period, as were average share values. The shares of total industry shipments among all large plants in industry 23 attributable to large plants in industry 232 were consistently the largest among 3-digit groups, ranging between 30 and 35 percent; the shares attributable to industry 239 were consistently second-largest, ranging between 20 and 25 percent. The shares attributed to industry 233 increased slightly during the period 1979-1986 but remained within

[27] All share values are not printed for all years: when two or more values overlap, only one is printed. For example, all actual share values obtained for industry 23 are provided in the following table; the shares do not sum to 100 due to rounding.

| | 231 | 232 | 233 | 234 | 235 | 236 | 238 | 239 |
|---|---|---|---|---|---|---|---|---|
| **1972** | 14 | 34 | 16 | 08 | 00 | 04 | 02 | 20 |
| **1973** | 14 | 32 | 16 | 08 | 00 | 04 | 02 | 22 |
| **1974** | 14 | 34 | 16 | 08 | 00 | 04 | 04 | 20 |
| **1975** | 12 | 36 | 16 | 08 | 00 | 04 | 02 | 22 |
| **1976** | 12 | 34 | 16 | 08 | 00 | 04 | 02 | 24 |
| **1977** | 10 | 36 | 16 | 08 | 00 | 04 | 02 | 24 |
| **1978** | 10 | 34 | 16 | 08 | 00 | 04 | 02 | 26 |
| **1979** | 10 | 36 | 18 | 08 | 00 | 04 | 02 | 22 |
| **1980** | 10 | 38 | 18 | 08 | 00 | 04 | 02 | 18 |
| **1981** | 10 | 36 | 20 | 08 | 02 | 04 | 02 | 20 |
| **1982** | 08 | 36 | 20 | 08 | 00 | 04 | 02 | 20 |
| **1983** | 08 | 34 | 18 | 08 | 00 | 04 | 02 | 24 |
| **1984** | 08 | 34 | 18 | 08 | 00 | 04 | 02 | 24 |
| **1985** | 10 | 32 | 16 | 08 | 00 | 06 | 02 | 26 |
| **1986** | 10 | 32 | 18 | 06 | 00 | 04 | 02 | 26 |

the range of approximately 15-25 percent over the entire period. The shares attributed to industry 231 declined from approximately 15 to approximately 10 percent. The share rankings for these two groups, however, remained stable after 1974, when they began to diverge from equal values. Shares for industry 234, 236 and 238 ranged between 3 and 8 percent, and were also consistent in their relative rankings. The smallest 3-digit shares reported for this group, for industry 235, ranged between 0 and 2 percent.

Although the plots varied from industry to industry, one general pattern is evident when reviewing these charts at the 2-digit level. It is that large plants in a single sub-group were responsible for an extremely large share of large plants' total output in only three of the non-

---

[28] All share values are not printed for all years: when two or more values overlap, only one is printed. For example, all actual share values obtained for industry 23 are provided in the following table; the shares do not sum to 100 due to rounding.

| | 231 | 232 | 233 | 234 | 235 | 236 | 238 | 239 |
|---|---|---|---|---|---|---|---|---|
| **1972** | 14 | 34 | 16 | 08 | 00 | 04 | 02 | 20 |
| **1973** | 14 | 32 | 16 | 08 | 00 | 04 | 02 | 22 |
| **1974** | 14 | 34 | 16 | 08 | 00 | 04 | 04 | 20 |
| **1975** | 12 | 36 | 16 | 08 | 00 | 04 | 02 | 22 |
| **1976** | 12 | 34 | 16 | 08 | 00 | 04 | 02 | 24 |
| **1977** | 10 | 36 | 16 | 08 | 00 | 04 | 02 | 24 |
| **1978** | 10 | 34 | 16 | 08 | 00 | 04 | 02 | 26 |
| **1979** | 10 | 36 | 18 | 08 | 00 | 04 | 02 | 22 |
| **1980** | 10 | 38 | 18 | 08 | 00 | 04 | 02 | 18 |
| **1981** | 10 | 36 | 20 | 08 | 02 | 04 | 02 | 20 |
| **1982** | 08 | 36 | 20 | 08 | 00 | 04 | 02 | 20 |
| **1983** | 08 | 34 | 18 | 08 | 00 | 04 | 02 | 24 |
| **1984** | 08 | 34 | 18 | 08 | 00 | 04 | 02 | 24 |
| **1985** | 10 | 32 | 16 | 08 | 00 | 06 | 02 | 26 |
| **1986** | 10 | 32 | 18 | 06 | 00 | 04 | 02 | 26 |

durable manufacturing groups, while in durable manufacturing a single sub-group, or two sub-groups that appeared to be "entwined," were responsible for an distinctly larger share of total large plants' shipments in most 2-digit groups. The three exceptions to this statement, among non-durable manufacturing establishments, were industries 21, 25, and 29. In industry 21, 75-90% of all shipments from large plants were attributed to large plants with a 3-digit SIC code of 211. Similarly, among large plants in industry 25, those in sub-group 251 accounted for 50-70% of the total value of shipments. Within industry 29, industry 291 was responsible for approximately 95% of the total value of shipments from large plants. For all other non-durable manufacturing groups (i.e., for industries 22, 23, 24, 26, 27 and 28), no single 3-digit sub-group exhibited a markedly dominant share of the total value of large plants' shipments.

In contrast, within durable manufacturing, in most cases one or two sub-groups were responsible for very large shares of the total value of shipments from large plants. For example, sub-groups 301 and 307 were jointly responsible for roughly 70% of the total value of shipments from large plants in industry 30, and roughly 75% of the total value of shipments from large plants in industry 31 was attributed to large plants with SIC code 314. The three exceptions to this general statement, among durable manufacturing establishments, were industries 34, 35, and, to a lesser degree, industry 38. In the case of the latter 2-digit groups, no sub-group share was dramatically larger than the others were, and relative rankings were stable.

### 2. Four-digit shares of shipments from large plants, by 3-digit group

Shipments shares of large plants in the 4-digit sub-groups within each 3-digit group of large plants tend to display greater variability in value and ranking than the 3-digit sub-group shares discussed in section II.B.1. Figures 5 and 6 illustrate the greater instability of sub-group shares of total shipments attributed to large plants in 3-digit industry groups. This instability appears to be characteristic of this lower level of aggregation. Figure 5 shows sub-group share plots for total shipments from large plants in industry 234. Over the period 1972-1978, sub-group shares for industry 2341 and 2342 were reasonably stable; these two groups accounted for slightly more than 60 percent and slightly less than 40 percent, respectively, of total shipments from large plants in industry 234. Over the period 1979-1985, the sub-group shares of industry 2341 increased to more than 70 percent, while the sub-group shares of industry 2342 declined to less than 30 percent. In 1986 both sub-group shares reverted to values similar to those observed during the earlier period. The relative rankings of the two sub-groups remained constant.

In Figure 6 we see greater instability of sub-group shares, both with respect to their values over time and with respect to their relative rankings, in the case of industry 238. The fluctuations of sub-group 2381 are especially dramatic, spanning high values of approximately 35 percent in 1973 and 1978, when industry 2381 was responsible for the largest shares of large plant' shipments among all sub-groups in industry 238, and declining to less than 10 percent and the smallest shares among all sub-groups in 1985 and 1986.

### C. Selection of 2-digit industry groups for further examination.

Based on our review of the share plots discussed in this Section, we narrowed the set of industries under consideration for detailed work to those with SIC codes 26, 28, 34, 35, and 38. The stability of large plants' shares of total industry output for these 2-digit groups indicated that large plants in these industries might reasonably be examined as a distinct subset, allowing us to defer analysis of the long-run plant size decision for future research. Similarly, the stability of sub-industry shares of large plants' industry shipments indicated that it would probably be acceptable to develop parameter estimates for systems of factor demand equations without incorporating a primary product switching equation. Information on average imputation rates by 4-digit industry and counts of observations available annually within 3-digit groups, discussed in Section III, were subsequently used to narrow the list of industries to be examined from 5 to 3.

## III. Imputation Rates and Annual Observations in Selected Industries

We checked on the imputation rates for the 4-digit groups included in industries 26, 28, 34, 35, and 36, using Attachment 5 of Roberts and Monahan's paper on nonsampling errors.[29] In this Attachment the value-weighted mean percentages of variables' industry-level values attributable to imputation are reported for each 4-digit industry group. These percentages range from a low of .01 (i.e. little imputation) for SIC code 2111 to a high of 48.82 (i.e., extensive imputation) for industry 3572. Although these percentages are reported for the LRD as a whole, they should be reasonably representative for large plants, given the magnitudes of large plants' contributions to the total value of shipments in manufacturing industries. Results from Roberts and Monahan's Attachment 5 are rearranged and reported in Tables 5.1-5.4.

After reviewing the share plots discussed in Section II and the imputation rates summarized in Tables 5.1-5.4, we made a preliminary decision to focus our analysis on large plants in industries 35 and 38. The following factors contributed to this decision. First, the shares of large plants' shipments in total industry shipments and the sub-industry shares among large plants are reasonably stable over the period examined for these two industries. As noted above, this finding indicates that that the production possibility frontiers for these industries may be reasonably stable, and that the estimation of systems of factor demand equations might be undertaken without explicit modeling of firm-level decisions regarding establishments' size and primary products. That is, this preliminary examination of these data suggested that they could be used to analyze establishment-level productivity and demand for factors of production despite the fact that firm-level data on each establishment were not available.[30]

Second, imputation rates fall below the 10% range in 54 of the 56 4-digit groups included within industries 35 and 38.[31] Since the imputation rates in Tables 5.1-5.4 are based on observations from establishments in all employment-size classes, and since large plants' variable values tend

---

[29] Mark Roberts and James Monahan (1986), cited in footnote 10 above, p. 146.

[30] This comment should not be taken to imply that the incorporation of firm-level microdata with the establishment-level microdata then available in the LRD would not be a desirable undertaking. Indeed, some of our analytical work has been devoted to identifying the simplifying assumptions that are imposed by necessity in order to analyze establishment-level productivity and demand for factors of production without reference to firm-level variables. But given that firm-level microdata were not then available, the share plots described here were helpful for selecting a small number of industries for detailed econometric analysis.

[31] The exceptionally high imputation rate of 49% for industry 3572 would have been a cause for concern, but in fact our files contained no observations on large plants in industry 3572. It should be noted, however, that the shipments price index for computers, industry 357, was considered unreliable, and it was being revised during our tenure with the ASA/NSF/Census Research Program. As a result, we have chosen not to emphasize results from industry 357 when 3-digit results are reported.

to be imputed less frequently than average, the integrity of these microdata seemed acceptable for the estimation of fairly complicated systems of equations.

Third, the fact that the production of durable goods is known to be highly cyclical, in combination with the fact that large plants were responsible for a large share of the total market in these industries, suggests that they are well-suited for applications involving systems of factor demand equations derived from inter-temporal objective functions. Since longitudinal microdata are particularly well-suited for the estimation of dynamic models, it seemed to us that the strengths of the LRD might be illustrated particularly well in applications involving data from industries such as these, in which dynamic optimization is likely to be standard practice.

## IV. Stability of composition of large plants within industry groups.

In this section the frequency with which establishments' SIC codes change from year to year, and the durations of spells during which establishments' primary product codes fall within those for a given 2-digit group, are reported for industries 35 and 38. In combination with the results reported in Sections II and III, this description of the stability of the industrial composition of the population of establishments provides preliminary evidence on the likelihood that aggregation bias may be a problem in index-number and econometric measures constructed with pooled industry time series data.[32]

### A. Frequency of individual plants' switches among industry categories

To screen for stability in the industrial composition of establishments within each 2-digit group, we counted the number of plant-years in which an establishment's SIC code changed, and generated frequency distributions for the number of years in which establishments' SIC codes changed to that of another industry group. These frequency distributions are reported in Table 6. They show that 1.6% of the annual observations coded in industry 35 were from plants whose 2-digit industry code was not 35 in the preceding year. The corresponding percentage for annual observations in industry 38 was 2.9%. These results indicate that the industrial composition of the plants in these two industries was quite stable over this period, and that the calculation of year-to-year changes should be straightforward for the vast majority of observations.

[32] This concern also underlies the formal econometric tests for separability that were implemented in the later stages of the project. A helpful introduction to the problem of aggregation bias is available in Henri Theil (1965), Linear Aggregation of Economic Relations, Amsterdam: North-Holland.

Information on the specific 2-digit groups with which plants that switched into these industries had been associated during the preceding year is provided in Table 7. This table shows that the industry of origin in the prior year was fabricated metal, machinery, transportation equipment or fine instruments in the vast majority of year-to-year changes. That is, the majority of prior-year products were broadly similar to current-year products, even when there was a change in 2-digit industry code. These results show that, in very general terms at least, the composition of outputs produced by these plants did not exhibit extreme variability from year to year.

### B. Duration of spells in industry and size class, by year

Although most annual observations showed no change in SIC code from the prior year, Tables 8.1 and 8.2 show that only about half of the large plant observations available for industry 35 in 1986 were from plants that were coded within industry 35 for the entire period examined. This table shows that, despite the fact that most large plants' SIC codes do not change from year to year, there is a great deal of churning over the longer run, as some plants switch to other industry groups or "downsize," and others enter or re-enter the product markets and employment size class examined here. In 1986, for example, 68% of all large plants in industry 35 had operated continuously in that industry, continuously employing 250 or more employees, for 10 years or longer. Those that had operated continuously in industry 35, employing 250 or more employees for 5 years or less, represented 23% of the total number of large plants in industry 35.

### C. Documented reasons for gaps in coverage

We were advised by Industry Division specialists that some plant identification numbers had been miscoded in the early 1980's, with the result that the number of true gaps in coverage identified for this period might be overstated.[33] We investigated this problem for the microdata on large plants in industries 35, 36, and 38 by examining distributions of the coverage code variable (CC) over time, and over employment size intervals. Among all large plants in industries 35, 36, and 38 for which there was at least a one-year discontinuity in coverage, the coverage code variable was coded as zero for 89% of all plant-year observations. This percentage was fairly stable, ranging between 76% in 1979 and 93% in 1984.[34] After deleting observations for which the coverage code variable indicated that a discontinuity was genuine, the percentage of large plants with a gap in coverage for which the coverage code variable was zero was also 89%, and again reasonably stable over time.

[33] A corrected file of plant identification numbers has subsequently been compiled by staff members of the Census Bureau's Center for Economic Studies.

[34] Genuine discontinuities were defined as observations with coverage code values of 13, 21, 22, 23, 28, 29, 31, 32, 33, 35, 38, 51, 54, 58, 64, 72-86, and 90. These variables are defined in Ruggles and Ruggles (1984), in the section titled "Dictionary Documentation," pp. 120-123

Among large plants whose gaps in coverage were not explained by coverage codes, 54% had total employment levels of 250-299, as compared with 15% of the total population of large plants in these industries. In the case of all large plants these percentages were relatively stable over time, as shown in Table 9. But there is a definite cyclical component to the presence of gaps in coverage, with no coverage code assigned, among plants in the 250-299 employee size class. During the expansionary period of 1975-1979, plants in this size class accounted for more than 70% of the discontinuous observations with questionable coverage codes, while they accounted for only 40-45 percent during the recessions of the early 1980's.

It seems likely that some of these observations represent switches from smaller employment-size classes, and it was not clear how these legitimate cases might be distinguished from mis-coded plant identification numbers without scanning the entire LRD file for observations with permanent plant identifiers from the observations in question.[35] Non-zero coverage codes among large plants with questionable gaps in coverage were distributed sparsely across all years, and across all employment-size-class intervals; they were not limited to the 1980's. Furthermore, the total number of observations with gaps not explained by the coverage code was relatively small: 1,602, or 3.2% of all plant-year observations. Given the difficulty of

[35] CES staff members did scan the entire LRD file for data on the same plant when it had fallen into another employment size class for the subsequent research with these data that is reported in Dhrymes (1991) and Bartlesman and Dhrymes (1992).

discerning how mistaken plant identification numbers should be corrected, the difficulty of separating coding errors from legitimate new entrants to this employment size class, and the relatively small number of cases in which questionable gaps might reflect mis-coded plant identification numbers, we decided that it would be inefficient to devote additional time to this issue.

### D. Frequency Distributions of Annual Observations

We examined the number of observations available annually for each three-digit group in industries 35, 36, and 38, in order to identify those three-digit groups with a sufficient number of annual observations to support the development of successive annual cross-section estimates of the parameters of simple systems of factor share equations. These frequency distributions, reported in Table 10, led us to select industries 353, 356, 357, 362, 366, and 367 for more detailed analysis.

## V. Deflators

### A. Four-digit industry price indexes

The microdata in the LRD are reported in current dollars; time series of price deflators are required to convert these nominal variables to "real" values, which are considered more appropriate for the estimation of production functions and factor demand functions. A review of the methodologies used to construct the industry-specific deflators available for use in this study revealed four matters that may warrant comment. First, data used to construct price indexes may have been collected from firms or from establishments. The total value of the shipments from a firm or establishment may be attributed to the industry represented by its primary product code, even when the total output of the firm or establishment is highly diversified. Second, the structure of the sample from which, and the selection of products for which, detailed and specific price data are collected is not identical for all series. Third, one must choose between fixed-weight and implicit price deflators in the case of plant and equipment investment. And fourth, input-output (I-O) tables and capital flows tables (CFTs) based on Census-year detailed product and materials data are used to approximate the composition of materials and energy used in production, the composition of investment, and the composition of the capital stocks employed in an industry.[36] These issues are familiar to experienced users of CM and ASM data, but they are reviewed briefly in this section for completeness.

---

[36] The construction of capital flows tables, or CFTs, is described in Peter E. Coughlin and Interindustry Economics Division Staff (1980), "New Structures and Equipment by Using Industries," Survey of Current Business, July, pp. 45-54. See the August 1971 and September 1975 issues of the Survey of Current Business for discussions of the CFTs for 1963 and 1967, respectively.

Aggregated industry time series are available on either an "establishment basis" or a "firm basis," or both in several instances. The former are aggregates of data collected from plants, while the latter are aggregates of data collected from firms. For example, the Census Bureau releases detailed estimates of current- and constant-dollar investments in plant and equipment. Two sets of estimates are released in each case. The establishment-based series include data on 58 industry groups, while the firm-based series include data on 39 industry groups. Differences in the two series arise because a single firm may own establishments whose primary product is different from the product to which the largest share of the firm's total shipments is attributed. Therefore the composition of investments attributed to all firms in a given industry may be different from the composition of investments attributed to all establishments in that industry.[37] From a conceptual standpoint, the establishment-based series are more appropriate for work with establishment-level microdata.[38] However, at the time that we investigated this matter the establishment-based series were intermediate products in the construction of the constant-dollar

---

[37] For a description of the methodologies used to construct the BLS price indexes, see Lawrence Grace, "Real Output Measurement in the United States National Income and Product Accounts," in Readings in Concepts and Methods of National Income Statistics, 1976, pp. 216-290; "Producer Prices," in The BLS Handbook of Methods, 1982, vol. 1., chapter 7, pp. 43-61; and "Wholesale Prices," in The BLS Handbook of Methods for Surveys and Studies, 1976, chapter 16, pp. 123-126. For a description of the methodology used to construct the constant-dollar series on plant and equipment investment, see Michael J. McKelvey (1981), "Constant Dollar Estimates of New Plant and Equipment Expenditures in the United States, 1947-80," Survey of Current Business, September, pp. 26-41; and Eugene P. Seskin and David F. Sullivan (1985), "Revised Estimates of New Plant and Equipment Expenditures in the United States, 1947-83," Survey of Current Business, January, pp. 16-25.

[38] We are grateful to John Gates for providing us with the references that describe the construction of these indexes, and for pointing out the possible usefulness of the establishment-based series.

firm-based investment series and were not subject to analyst review for "sensibleness." Since the firm-based series was subject to analyst review, and since the data were originally collected at the firm level, we decided to work with the more familiar firm-based series.

The sample of establishments from which specific price data are collected for construction of the BLS Producer Price Index (PPI) is designed to monitor the price changes of those particular products within an industry group that are responsible for the largest values of shipments. Large plants represent a disproportionate share of the establishments from which these data are collected. More randomness has been introduced into the selection of plants whose specific product prices are used to construct the Industry Price Index, (IPI), which measures changes in the prices of all products produced by the firms or establishments within a given industry, even when some of the products produced are themselves classified in other industry categories.

The Census-SRI-Penn-NBER (CSPN) dataset of industry deflators and aggregates, which was readily available for our use, is based on PPIs through 1976; IPIs were used as they became available, over the period 1977-1984.[39] The deflators of choice for the purposes of this

[39] For a description of the procedures followed to construct the original Census-SRI-Penn dataset through 1976, see Gary Fromm, David L. Crawford, Lawrence R. Klein, and Frank C. Ripley, "The Census-SRI-Penn Industry Profiles Dataset," unpublished report, June 15, 1982; also see Steve Andrews and Craig Zabala, "Documentation of the SRI-Penn Manufacturing Industry Dataset Developed by David L. Crawford, Gary Fromm, Lawrence Klein and Frank C. Ripley, "Bureau of the Census Technical Notes, 1984. A description of

project might have been consistent series based on the PPI, had they been available without requiring extensive re-coding and re-aggregation, because the population from which these price data are taken corresponds most closely to the population under study. However, the time and resources that would have been required to duplicate the construction of materials deflators from these shipments price indexes, following the procedures described by Fromm, Klein et. al., would have been prohibitive. Furthermore, if a project were to be undertaken to construct materials deflators for use with the LRD files, it would probably be more worthwhile to construct plant-specific deflators, using detailed product and materials data from Census years, than it would be to duplicate the CSPN procedures using PPIs.[40]

Both fixed-weight deflators and implicit price deflators are available separately for plant investment and equipment investment. The differences between the methodologies used to construct the two types of deflators are well known. The former can be inaccurate measures of changes in the costs that plants actually face, because the composition of investments purchased will change in response to changes in the relative prices of capital goods and other factors of production, and in response to changes in the composition of aggregate demand for products

---

the procedures followed to update this dataset through 1984 was provided in Wayne B. Gray (1987), "Productivity Data Description," photocopy.

[40] We owe this recommendation to James Monahan of the Census' Center for Economic Studies.

produced.  On the other hand, price changes and compositional changes are confounded in implicit price deflators.

Industry-specific deflators for plant and equipment investments that are disaggregated beyond the two-digit level were not available for years after 1981 during our tenure in the ASA/NSF/Census Research Program.   Consequently we decided to forecast the industry deflators after 1981 as functions of plant and equipment deflators for all manufacturing, which were available after 1981 in the NIPA Tables 7.4, 7.12, and 7.13.  We found that the equations forecasting the industry-specific implicit price deflators for equipment had a significantly better fit than the corresponding equation for the fixed-weight deflators, and we chose to work with the implicit price deflators for this reason.

### B. Establishment-specific investment deflators

We constructed plant-specific deflators for total investments, weighting the firm-based implicit price deflators discussed above by the shares of total investment attributed to machinery and building investments, at the establishment level.    That is, industry price deflators for plant and equipment, denoted $P_{I,p}$ and $P_{I,e}$ respectively, were merged with establishment-level data on current-year expenditures on plant and equipment. Plant-specific investment deflators $P_i$ were calculated according to the following formula:

$$P_i = \left( \frac{NB_i}{NB_i + NM_i} \right) P_{I,P} + \left( \frac{NM_i}{NB_i + NM_i} \right) P_{I,E} ,$$

where $NB_i$ and $NM_i$ represent new building expenditure and new machinery expenditure respectively.

The resulting time series were extremely variable, due to the variability of building investments at the establishment level. Consequently this approach to the construction of establishment-level investment deflators did not seem to result in a single index that measured the prices of investment goods actually faced by each establishment on an annual basis. We therefore decided to deflate investments in machinery and plant separately by the industry-level implicit price deflators for plant and equipment, described above. We used the resulting deflated investment values to construct separate estimates of establishment-level stocks of plant and equipment according to the perpetual inventory approach, and aggregated the two capital stock estimates to obtain an estimate of the total constant-dollar value of plant and equipment held by each establishment.

## C. Calculation of value added.

In their documentation of the LRD Ruggles and Ruggles note that there are several differences in definition and measurement between the Census Bureau's measure of value added, which is available in these microdata, and the Bureau of Economic Activity's (BEA's) output measure, gross product originating in manufacturing (GPO).[41] Both measures begin by measuring output as the total value of shipments (TVS) plus net changes in inventory of work in progress (WIE-WIB) and inventory of finished products (FIB-FIE) minus the cost of materials (CM). But the following differences in definition are noted: VA includes the cost of purchased services while GPO does not; GPO includes an estimate of the value of taxes other than property taxes while VA does not, and GPO includes an adjustment that is intended to convert the reported value of inventories to a replacement cost basis.[42] To make an initial check on the internal consistency of the microdata we calculated VA directly with reported values for TVS, WIE, WIB, FIB, FIE and CM, and compared the directly-calculated value with the value of VA recorded in the

---

[41] Ruggles and Ruggles (1984), cited above, pp. 104-107. The Ruggles cite the 1982 Census General Summary Part 1 (Census Report number MC82-S Part 1) as a reference for more detailed comparisons between Census' VA and BEA's GPO.

[42] The Ruggles' also note that the information sources used to construct VA and GPO differ in several respects, and that "GPO has tended to be 10 to 20 percent lower than value added" as a result of these differences in definition and data sources.

LRD. An extremely small number of discrepancies was discovered, and we concluded that we need have no concern with the internal consistency of the recorded values.[43]

## VI. "Perpetual inventory" capital stock estimates.

### A. Construction of Capital Stock Estimates

We estimated the real value of the stocks of plant and equipment available for production according to the "perpetual inventory technique." Although the LRD includes information on the book value of buildings and machinery, these numbers are generally considered to be unreliable measures of the true productive capacity of the capital stock because they are not adjusted to account for changes in nominal value over time.[44]

The perpetual inventory technique is employed to construct industry-wide time series of capital stock estimates. It is a procedure in which constant-dollar values of industry investments are

[43] In fact, the number of discrepancies discovered was so small that the output from these checks was not retained when the term of our formal tenure in the ASA/NSF/Census Research program had expired. However, our memory is that the number of observations with such discrepancies was fewer than 10.

[44] A lucid discussion of the reasons why capital stock estimates constructed according to the perpetual inventory technique are preferred over book value measures is provided in "Capital Stock Estimates for Input-Output Industries: Methods and Data," Bureau of Labor Statistics Bulletin 2034, 1979.

cumulated, after adjustment to account for declines in the efficiency of productive physical assets as they age. This adjustment is based on the following relationship:

$$K_{I,t}^{g} = \left(e - d_{I}^{g}\right)G_{I,t}^{g} = s_{I}^{g} G_{I,t}^{g},$$

where $e = (1,1,\ldots 1)$. The share of investments in assets of type $g$ made at time $t-n$ that become unproductive at time $t$ is $d_{I,t-n}^{g}$, so $s_{I,t-n}^{g}$ is the surviving fraction of productive physical assets purchased at time $t-n$, as of time $t$. The entire vector of efficiency adjustments, $d_{I}^{g} = (d_{I,t}^{g}, d_{I,t-1}^{g} \ldots, d_{I,t-z}^{g})$, is defined so that $d_{I,t}^{g} + d_{I,t-1}^{g} + \cdots + d_{I,t-l}^{g} = 1$, where $l$ is the maximum service life of the asset, i.e. $d_{I,t-l-k}^{g} = 0$ for all $k > 0$. The vector of investments is $G_{I,t}^{g} = (I_{I,t}^{g}, I_{I,t-1}^{g}, \ldots I_{I,t-z}^{g})'$. $I_{I,t-n}^{g}$ is gross investment in assets of type $g$ made at time $t-n$ by establishments in industry $I$, measured in constant dollars. The elements of $s_{I}^{g}$, $s_{t-n}^{g}$, are functions of the expected duration and variance of asset service lives, as well as the specific hyperbolic function that is assumed to characterize the relationship between age and efficiency among assets of a specified type.[45]

[45] A more detailed description of these calculations is provided in BLS Bulletin 2034, cited above.

Time limitations prevented the full implementation of this approach with the microdata in the LRD. Rather, the elements of the vector $s_I^g$ were all assumed to take the value $s$, as estimated from the following regression, $K_{I,t} - I_{I,t} = \left(1-\tilde{d}\right)_I K_{I,t-1} = \tilde{s}_I K_{I,t-1}$ where $K_{I,t,}K_{I,t-1}$ and $I_{I,t}$ are industry time series produced by the Bureau of Industrial Economics, and $K_{I,t} = \sum_g K_{I,t}^g$.

The BIE time series are produced by applying the perpetual inventory approach to an establishment-based time series on industry investments at the three-digit level that dates back to 1958. Because the useful life of assets purchased in 1958 is assumed to have ended by 1972, the initial conditions assumed by BIE in the construction of these time series are largely immaterial, for the purposes at hand. However the initial conditions used to construct plant-level capital stock estimates are significant, especially in the early years spanned by the microdata.

Our procedure was as follows. We estimated each plant's initial capital stock as a share of the industry-wide estimate:

$K_{i,0} = \left(S_{i,0}/S_{I,0}\right)K_{I,0}$, where letters subscripted $i$ denote plant-level data, letters subscripted $I$ denote aggregate time series values for the industry in which the plant's primary product is categorized in year $0$, and the subscript $0$ corresponds to the first year in which an observation on plant $i$ is available. The symbols $K_{i,0}$ and $K_{I,0}$ thus represent plant-level and industry-level capital stock estimates, and $S_{i,0}$ and $S_{I,0}$ represent plant-level and industry-level values of annual shipments.

Since the tapes from which the microdata under study were extracted contain observations on all plants with 250 or more employees it was possible to obtain investment data on plants whose primary product code switched from 35, 36 or 38 to some other value, and this was done. However time limitations prevented the correction of cases in which plant identification numbers had been miscoded, and the extraction of investment data on plants whose employment levels declined below 250. These cases were treated as gaps in coverage, and assigned new initial values. For each initial value for $K_{i,0}$, plant-level capital stock estimates for years $t > 0$ were then constructed as follows.

$$K_{i,t} = I_{i,t} + (1 - \tilde{d}_I) K_{i,,t-1.} = I_{i,t} + \tilde{s}_I K_{i,,t-1.}$$

### B. Distributions of Capital Stock Estimates

Distributions of total estimated capital stock were compared with those for total employment, value added, and the total value of shipments. These distributions are reported in Tables 11-14. They are extremely skewed, as expected for certainty cases from an optimally stratified sample.

## VII. Summary and Conclusion

This paper has reviewed results obtained in a preliminary investigation of the Census Bureau's Longitudinal Research Data file. The objective of this preliminary research was to select a subset of LRD establishment-level observations that would support the substantive objectives of our research project.

In our review of documentation developed by Ruggles and Ruggles we learned that imputation rates are lowest, and continuity of coverage is greatest, among observations on large plants that are sampled with certainty in the ASM. Examining the share of total industry shipments attributable to large plants at the 3- and 4-digit level, we found that large-plant shares of total shipments by 3-digit industry exhibited significantly greater stability than large-plant shares within 4-digit groups. Large-plant shares of total industry shipments in industries 35, 36 and 38 were distinctly larger than the average for all manufacturing plants.

Examining the rankings of sub-industry shares of all large plants' shipments at the 2-digit level, we found that these shares are dominated by one or two sub-industries in most of durable manufacturing; this pattern is observed much less frequently in the nondurable manufacturing industries. Among large plants producing durable goods, sub-industry shares were most evenly distributed within industries 34, 35, and to a lesser extent, 38. We again found greater fluctuations among 4-digit shares of 3-digit large plant totals, compared with 3-digit shares of 2-digit large plant totals.

Review of published value-weighted imputation rates for all plants at the 4-digit industry level revealed that imputation rates are consistently below 10% among the industries we selected for analysis. Since imputation rates are significantly lower among large plants than for the industry as a whole, these published rates should provide upper bounds on the true imputation rates for large plant data.

Pooling observations for the period 1973-1986, we found that the frequency with which plants' primary products in the prior year fell within a different 4-digit industry group than that reported for the current year was fairly low, i.e., less than 3%. Over the longer run, however, the degree of sustained industry attachment was substantially lower. Examination of coverage codes for large plants with gaps in coverage seemed to indicate that the percentage of all plant-year observations with gaps that might be attributable to errors in the assignment of permanent plant identification numbers was relatively small. A routine procedure for the identification and correction of incorrect identification numbers was not evident, and none was implemented. However CES staff members have subsequently developed a master list of corrected permanent plant identification numbers that should be useful in the future.

We devoted some effort to the development of establishment-specific investment deflators, but found that the resulting plant-level deflators were quite unstable due to the "lumpiness" of building investments. However the calculations required to construct plant-level deflators proved to be relatively straightforward. Construction of plant-level capital stocks through the perpetual inventory technique yielded estimates whose distributions are comparable to those obtained with directly reported data on employment and the total value of shipments.

In summary, we believe the time-series cross-section microdata in the LRD to be an important new source of economic information. The variables available in the LRD are described extensively in documentation assembled initially by Ruggles and Ruggles. In additional useful documentation has been developed by members of the CES staff, and by outside researchers. The LRD is an extremely large file, but preliminary review of selected subsets of observations, for a small subset of key variables, can be used constructively to select of subset of observations that is well-suited for the research objectives at hand.

Table 1
LRD Estimates for the Total Value of Shipments
Totals for Plants Linked in Successive Years
Relative to Size Class Totals

| 1-99 | 100-249 | 250-499 | 500+ | overall |
|---|---|---|---|---|
| 18.8 | 53.3 | 82.4 | 91.41 | 69.9 |

Table 2
LRD Estimates for the Total Value of Shipments
Reported Totals Relative to Reported and Imputed Totals

| Employment size class | 1974 | 1981 |
|---|---|---|
| 1-99 employees | 74.1 | 88.2 |
| 100-249 employees | 81.5 | 91.4 |
| 250-499 employees | 90.4 | 94.5 |
| 500+ employees | 98.4 | 98.5 |
| overall | 91.6 | 96.0 |

Table 3
LRD Estimates for Selected Variables
Percentage Reported by Employment Size Class

| | | Salaries & Wages | | Shipments | | Other, Average | |
|---|---|---|---|---|---|---|---|
| Year | Emp's | reported | $ value | reported | $ value | reported | $ value |
| 1972 | all | 99% | 99% | 83% | 98% | 48% | 95% |
| | 0-99 | 99 | 99 | 82 | 95 | 42 | 80 |
| | 100-249 | 98 | 98 | 97 | 98 | 93 | 96 |
| | 250-499 | 99 | 99 | 98 | 99 | 97 | 98 |
| | 500+ | 100 | 100 | 100 | 100 | 99 | 100 |
| 1973 | all | 92% | 97% | 87% | 97% | 88% | 96% |
| | 0-99 | 91 | 93 | 84 | 91 | 85 | 90 |
| | 100-249 | 93 | 94 | 90 | 94 | 90 | 94 |
| | 250-499 | 95 | 96 | 93 | 96 | 93 | 95 |
| | 500+ | 98 | 97 | 97 | 99 | 97 | 99 |
| 1974 | all | 96% | 98% | 68% | 91% | 89% | 97% |
| | 0-99 | 96 | 96 | 61 | 74 | 87 | 92 |
| | 100-249 | 96 | 96 | 70 | 81 | 92 | 95 |
| | 250-499 | 96 | 96 | 81 | 90 | 94 | 96 |
| | 500+ | 97 | 99 | 95 | 98 | 97 | 99 |
| 1975 | all | 91% | 96% | 89% | 97% | 89% | 97% |
| | 0-99 | 89 | 89 | 86 | 92 | 87 | 92 |
| | 100-249 | 92 | 93 | 92 | 94 | 92 | 94 |
| | 250-499 | 94 | 95 | 94 | 96 | 94 | 96 |
| | 500+ | 98 | 99 | 98 | 99 | 98 | 99 |
| 1976 | all | 94% | 97% | 88% | 97% | 89% | 96% |
| | 0-99 | 94 | 93 | 84 | 91 | 85 | 92 |
| | 100-249 | 94 | 94 | 91 | 93 | 92 | 93 |
| | 250-499 | 95 | 95 | 94 | 96 | 94 | 95 |
| | 500+ | 97 | 99 | 97 | 99 | 97 | 981 |
| 1977 | all | 100% | 100% | 64% | 95% | 47% | 94% |
| | 0-99 | 100 | 100 | 60 | 84 | 41 | 78 |
| | 100-249 | 100 | 100 | 92 | 95 | 92 | 94 |
| | 250-499 | 100 | 100 | 96 | 97 | 96 | 97 |
| | 500+ | 100 | 100 | 98 | 100 | 98 | 100 |

Table 3, Continued
LRD Estimates for Selected Variables
Percentage Reported by Employment Size Class

| Year | Emp's | Salaries & Wages reported | Salaries & Wages $ value | Shipments reported | Shipments $ value | Other, Average reported | Other, Average $ value |
|---|---|---|---|---|---|---|---|
| 1978 | all | 94% | 98% | 86% | 97% | 87% | 97% |
| | 0-99 | 94 | 95 | 81 | 90 | 82 | 91 |
| | 100-249 | 95 | 95 | 91 | 93 | 91 | 95 |
| | 250-499 | 96 | 97 | 94 | 96 | 94 | 96 |
| | 500+ | 98 | 99 | 97 | 99 | 98 | 99 |
| 1979 | all | 93% | 98% | 84% | 96% | 84% | 97% |
| | 0-99 | 92 | 94 | 78 | 85 | 78 | 85 |
| | 100-249 | 93 | 94 | 87 | 91 | 87 | 91 |
| | 250-499 | 95 | 96 | 92 | 95 | 93 | 95 |
| | 500+ | 98 | 99 | 97 | 99 | 97 | 99 |
| 1980 | all | 92% | 97% | 83% | 96% | 84% | 96% |
| | 0-99 | 91 | 92 | 76 | 86 | 77 | 86 |
| | 100-249 | 93 | 93 | 86 | 91 | 87 | 91 |
| | 250-500 | 94 | 95 | 91 | 94 | 91 | 95 |
| | 500+ | 97 | 98 | 96 | 99 | 96 | 99 |
| 1981 | all | 100% | 100% | 84% | 96% | 85% | 95% |
| | 0-99 | | | | | | |
| | 100-249 | | | | | | |
| | 250-500 | 100 | 100 | 91 | 95 | 91 | 94 |
| | 500+ | 100 | 100 | 96 | 99 | 96 | 99 |

Table 4
LRD Observations on Establishments
With 250 or More Employees

| year | plants | % of total |
|---|---|---|
| 1972 | 13,590 | 6.8 |
| 1973 | 14,218 | 7.1 |
| 1974 | 14,241 | 7.1 |
| 1975 | 12,681 | 6.3 |
| 1976 | 13,211 | 6.6 |
| 1977 | 14,092 | 7.1 |
| 1978 | 14,051 | 7.0 |
| 1979 | 14,384 | 7.2 |
| 1980 | 13,766 | 6.9 |
| 1981 | 13,311 | 6.7 |
| 1982 | 13,067 | 6.5 |
| 1983 | 12,098 | 6.1 |
| 1984 | 12,778 | 6.4 |
| 1985 | 12,404 | 6.2 |
| 1986 | 11,981 | 6.0 |
| 1972-1986 | 199,873 | 100% |

Figure 1: Large plants' shares of total industry shipments
SIC 20: Food and kindred products

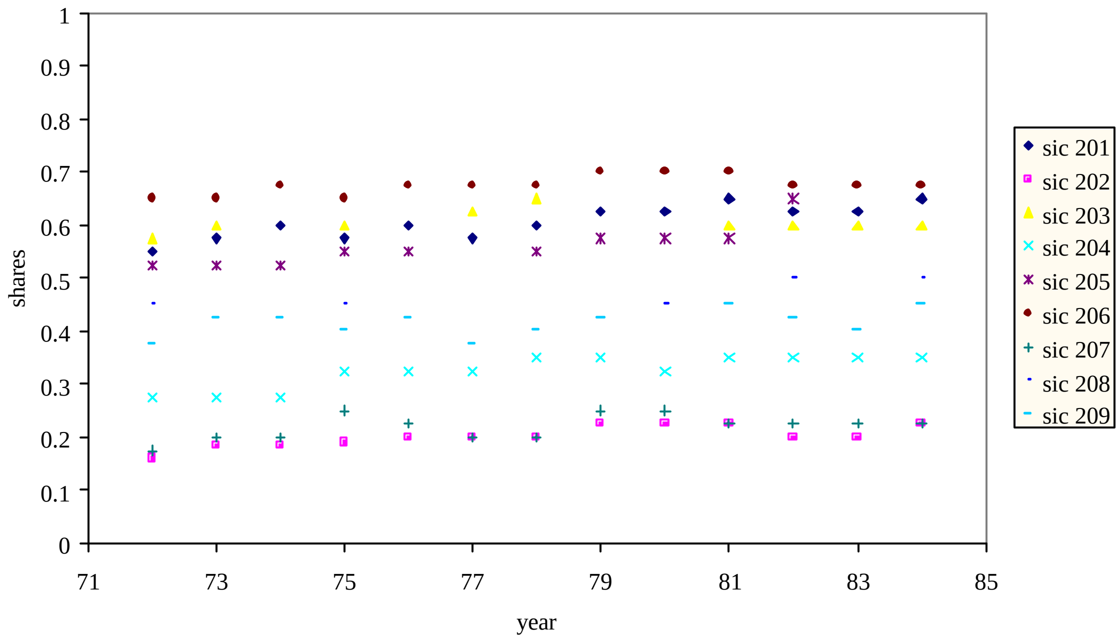

Figure 2: Large plants' shares of total industry shipments
SIC 234: Women's and children's undergarments

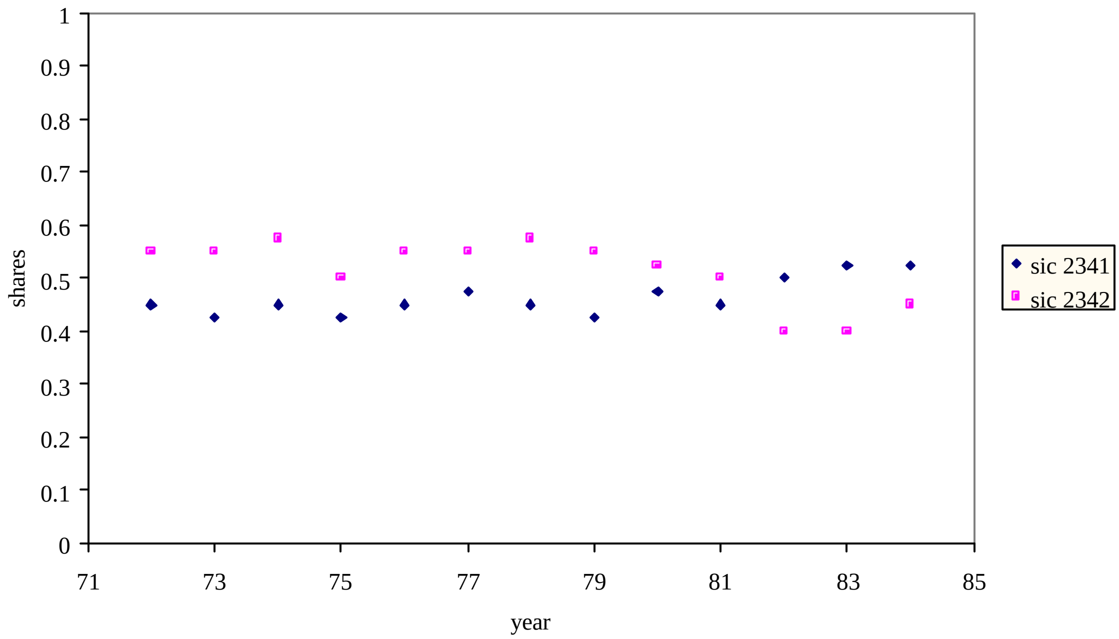

Figure 3:  Large plants' shares of total industry shipments
SIC 238:  Miscellaneous apparel and accessories

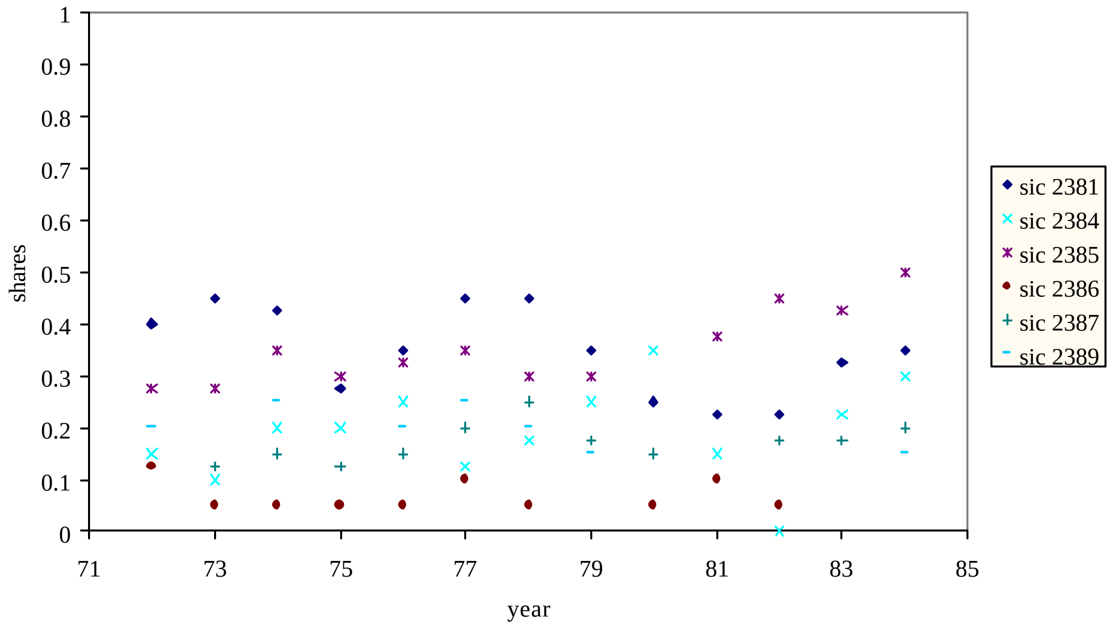

Figure 4: Sub-industry shares of shipments from large plants
SIC 23: Apparel and other textile products

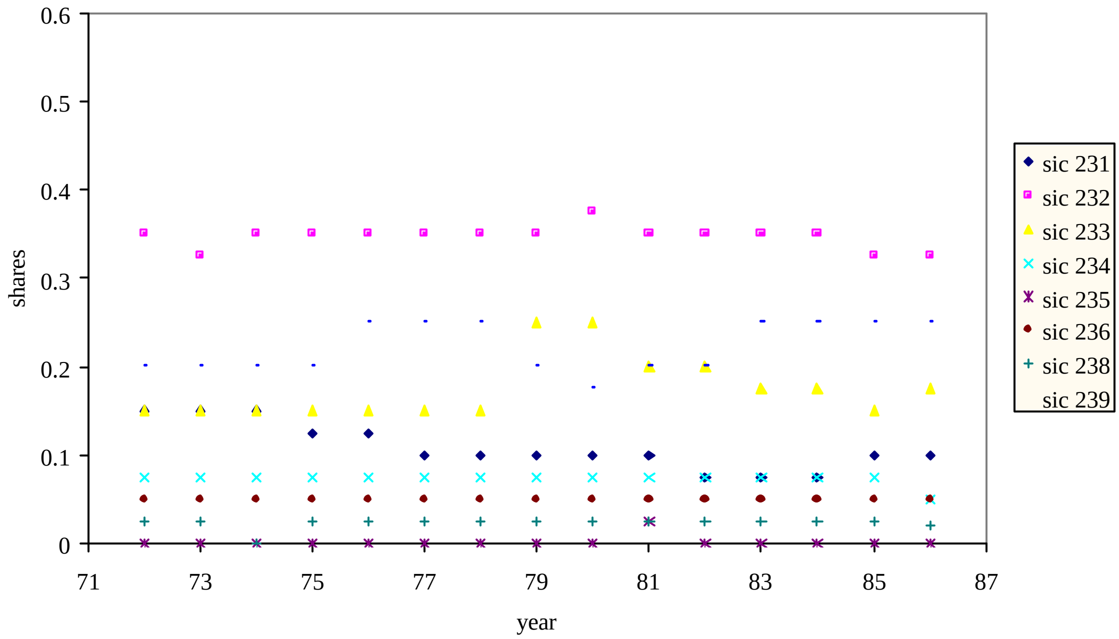

Figure 5: Sub-industry shares of shipments from large plants
SIC 234: Women's and children's undergarments

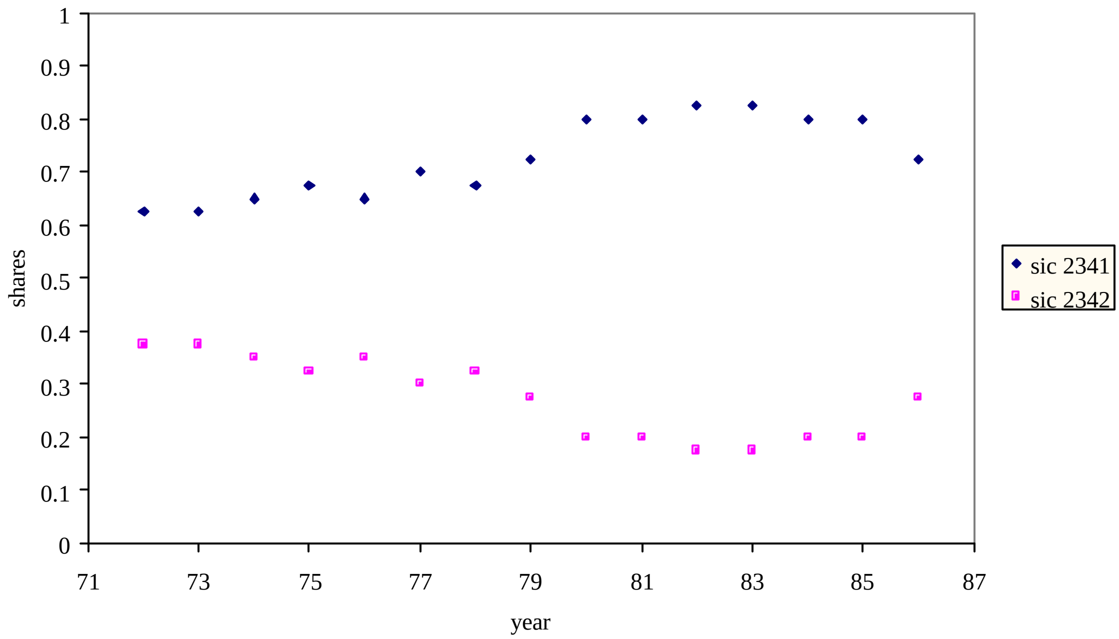

Figure 6:  Sub-industry shares of shipments from large plants
SIC 238:  Miscellaneous apparel and accessories

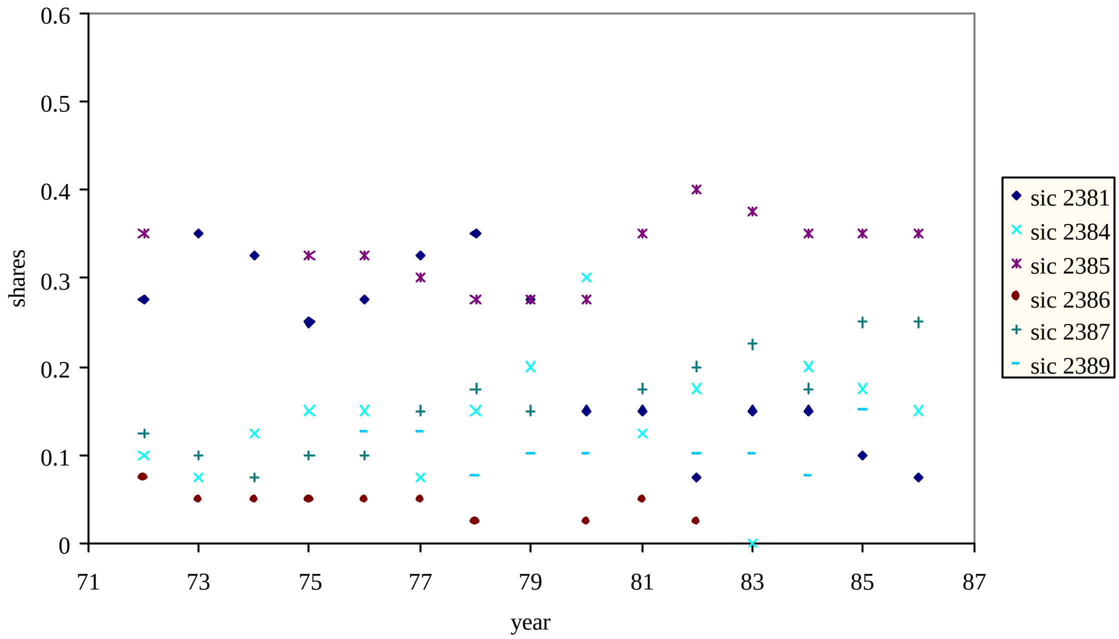

Table 5.1
LRD Imputation Rates
Value-weighted Mean Percentage Variables' Values
Attributable to Imputation[46]

| SIC 26 | % | SIC 28 | % | SIC 28 | % |
|---|---|---|---|---|---|
| 2611 | 0.22 | 2812 | 0.62 | 2861 | 6.45 |
| 2621 | 1.57 | 2813 | 5.69 | 2865 | 2.50 |
| 2631 | 3.43 | 2816 | 3.02 | 2869 | 1.79 |
| 2641 | 4.29 | 2819 | 1.93 | 2873 | 1.59 |
| 2642 | 5.69 | 2821 | 2.74 | 2874 | 1.67 |
| 2643 | 5.59 | 2822 | 0.47 | 2875 | 7.36 |
| 2645 | 8.99 | 2823 | 0.06 | 2879 | 4.54 |
| 2646 | 1.21 | 2824 | 0.45 | 2891 | 5.97 |
| 2647 | 2.46 | 2831 | 4.52 | 2892 | 5.08 |
| 2648 | 11.24 | 2833 | 2.25 | 2893 | 7.74 |
| 2649 | 7.41 | 2834 | 1.73 | 2895 | 0.30 |
| 2651 | 6.49 | 2841 | 2.05 | 2899 | 6.91 |
| 2652 | 16.42 | 2842 | 7.67 | | |
| 2654 | 4.27 | 2843 | 5.99 | | |
| 2655 | 7.91 | 2844 | 4.39 | | |
| 2661 | 1.02 | 2851 | 5.55 | | |

[46] Roberts and Monahan (1986), cited above.

Table 5.2
LRD Imputation Rates
Value-weighted Mean Percentage Variables' Values
Attributable to Imputation

| SIC 33 | % | SIC 34 | % | SIC 34 | % |
|---|---|---|---|---|---|
| 3312 | 1.07 | 3411 | 2.82 | 3482 | 1.44 |
| 3313 | 2.60 | 3412 | 5.83 | 3483 | 4.90 |
| 3315 | 6.43 | 3421 | 3.93 | 3484 | 4.05 |
| 3316 | 4.80 | 3423 | 4.26 | 3489 | 3.24 |
| 3317 | 3.70 | 3425 | 3.66 | 3493 | 2.96 |
| 3321 | 2.51 | 3429 | 3.48 | 3494 | 4.63 |
| 3322 | 1.15 | 3431 | 3.94 | 3495 | 6.89 |
| 3324 | 3.05 | 3432 | 5.47 | 3496 | 9.14 |
| 3325 | 3.38 | 3433 | 9.26 | 3497 | 4.20 |
| 3331 | 1.05 | 3439 | 12.79 | 3498 | 4.77 |
| 3332 | 0.02 | 3441 | 8.67 | 3499 | 9.71 |
| 3333 | 0.91 | 3442 | 13.15 | | |
| 3334 | 0.69 | 3443 | 5.85 | | |
| 3339 | 1.64 | 3444 | 9.58 | | |
| 3341 | 10.17 | 3446 | 6.77 | | |
| 3352 | 2.50 | 3448 | 8.46 | | |
| 3353 | 0.54 | 3440 | 5.94 | | |
| 3354 | 2.39 | 3451 | 8.82 | | |
| 3355 | 2.03 | 3452 | 4.39 | | |
| 3356 | 8.77 | 3462 | 2.84 | | |
| 3357 | 2.52 | 3463 | 1.00 | | |
| 3361 | 4.98 | 3465 | 1.60 | | |
| 3362 | 6.33 | 3466 | 0.91 | | |
| 3369 | 5.84 | 3469 | 6.59 | | |
| 3398 | 10.05 | 3471 | 12.62 | | |
| 3399 | 4.57 | 3479 | 7.49 | | |

Table 5.3
LRD Imputation Rates
Value-weighted Mean Percentage Variables' Values
Attributable to Imputation

| SIC 35 | % | SIC35 | % | SIC 36 | % |
|---|---|---|---|---|---|
| 3511 | 0.29 | 3559 | 4.76 | 3612 | 2.04 |
| 3519 | 0.37 | 3561 | 2.90 | 3613 | 2.49 |
| 3523 | 2.37 | 3562 | 1.05 | 3621 | 2.53 |
| 3524 | 2.97 | 3563 | 2.07 | 3622 | 3.26 |
| 3531 | 1.89 | 3564 | 7.50 | 3623 | 3.51 |
| 3532 | 2.90 | 3565 | 12.19 | 3624 | 0.80 |
| 3533 | 3.13 | 3566 | 2.45 | 3629 | 5.75 |
| 3534 | 3.80 | 3567 | 5.97 | 3631 | 2.14 |
| 3535 | 8.34 | 3568 | 1.57 | 3632 | 0.23 |
| 3536 | 2.83 | 3569 | 8.30 | 3633 | 0.38 |
| 3537 | 3.42 | 3572 | 48.82 | 3634 | 2.86 |
| 3541 | 2.64 | 3573 | 3.02 | 3635 | 1.88 |
| 3542 | 4.24 | 3574 | 1.49 | 3636 | 1.40 |
| 3544 | 9.13 | 3576 | 4.09 | 3639 | 2.91 |
| 3545 | 3.17 | 3579 | 1.89 | 3641 | 1.22 |
| 3546 | 1.05 | 3581 | 6.13 | 3643 | 3.46 |
| 3547 | 1.47 | 3582 | 6.19 | 3644 | 3.42 |
| 3551 | 7.16 | 3585 | 2.34 | 3645 | 10.95 |
| 3552 | 5.89 | 3586 | 3.01 | 3646 | 5.46 |
| 3553 | 5.38 | 3509 | 6.54 | 3647 | 1.27 |
| 3554 | 4.90 | 3592 | 0.80 | 3648 | 4.11 |
| 3555 | 7.05 | 3599 | 13.94 | 3651 | 3.93 |

Table 5.4
LRD Imputation Rates
Value-weighted Mean Percentage Variables' Values
Attributable to Imputation

| SIC 36 | % | SIC 38 | % |
|---|---|---|---|
| 3652 | 13.02 | 3811 | 5.29 |
| 3661 | 1.39 | 3822 | 1.26 |
| 3662 | 3.11 | 3823 | 3.26 |
| 3671 | 0.71 | 3524 | 1.75 |
| 3672 | 1.50 | 3825 | 4.42 |
| 3673 | 0.71 | 3829 | 6.17 |
| 3674 | 3.85 | 3832 | 5.39 |
| 3675 | 4.29 | 3841 | 4.12 |
| 3676 | 5.93 | 3842 | 4.16 |
| 3677 | 8.27 | 3843 | 4.33 |
| 3678 | 3.52 | 3851 | 6.46 |
| 3679 | 7.25 | 3861 | 1.72 |
| 3691 | 1.64 | 3873 | 5.01 |
| 3692 | 2.01 | | |
| 3693 | 3.77 | | |
| 3694 | 2.78 | | |
| 3699 | 10.01 | | |

Table 6
Annual Observations on Large Plants in the LED
Industries 35 and 38, 1973-86
Incidence of SIC Code Switches
X Denotes a Digit Switch
O Denotes No Change

| Switch Pattern | Industry 35 | | Industry 38 | |
|---|---|---|---|---|
| | N | % | N | % |
| OOOO | 18,343 | 96.9 | 5,511 | 95.0 |
| OXOO | * | 0.0 | * | 0.1 |
| OOXO | 40 | 0.2 | * | 0.1 |
| OOOX | 169 | 0.9 | 56 | 1.0 |
| XOXO | * | 0.0 | * | 0.0 |
| XOOX | | | * | 0.0 |
| OXXO | 31 | 0.2 | 18 | 0.3 |
| OXOX | 45 | 0.2 | 12 | 0.2 |
| OOXX | 134 | 0.7 | 58 | 1.0 |
| XXXO | | | * | 0.0 |
| XXOX | * | 0.0 | * | 0.1 |
| XOXX | | | * | 0.1 |
| OXXX | 200 | 1.1 | 107 | 1.8 |
| XXXX | * | 0.0 | 12 | 0.2 |

* Indicates 10 or fewer observations

Table 7
Annual Observations on Large Plants in the LED
Industries 35 and 38, 1973-1986
Industry in Previous Year

| Previous Industry | Industry 35 | | Industry 38 | |
|---|---|---|---|---|
| | N | % | N | % |
| Textile Mills | * | 0.0 | * | 0.1 |
| Apparel & Similar | | | * | 0.1 |
| Lumber & Wood | | | * | 0.0 |
| Furniture & Fixtures | * | 0.0 | * | 0.1 |
| Paper & Allies Products | * | 0.0 | * | 0.1 |
| Printing & Publishing | * | 0.0 | | |
| Chemicals & Allied | * | 0.0 | * | 0.1 |
| Rubber & Misc. Plastic | * | 0.0 | * | 0.1 |
| Leather | | | * | 0.0 |
| Stone, Clay, Glassware | * | 0.0 | * | 0.1 |
| Primary Metal | 16 | 0.1 | * | 0.0 |
| Fabricated Metal | 77 | 0.4 | 17 | 0.3 |
| Machinery Expt. Electrical | 18,686 | 98.4 | 26 | 0.4 |
| Electrical Machinery | 84 | 0.4 | 77 | 1.3 |
| Transportation Equipment | 60 | 0.3 | * | 0.1 |
| Fine Instruments | 29 | 0.2 | 5,632 | 97.1 |
| Miscellaneous | * | 0.0 | * | 0.1 |

* Indicates 10 or fewer observations

Table 8
Industry 35, 1973-1986

| Yr | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 72 | 1,322 | | | | | | | | | | | | | | |
| | 100.0 | | | | | | | | | | | | | | |
| 73 | 164 | 1,265 | | | | | | | | | | | | | |
| | 11.48 | 88.52 | | | | | | | | | | | | | |
| 74 | 160 | 141 | 1,226 | | | | | | | | | | | | |
| | 10.48 | 9.23 | 80.29 | | | | | | | | | | | | |
| 75 | 77 | 103 | 103 | 1,129 | | | | | | | | | | | |
| | 5.45 | 7.29 | 7.29 | 79.96 | | | | | | | | | | | |
| 76 | 68 | 64 | 95 | 103 | 1,074 | | | | | | | | | | |
| | 4.84 | 4.56 | 6.77 | 7.34 | 76.50 | | | | | | | | | | |
| 77 | 130 | 69 | 69 | 112 | 103 | 1,029 | | | | | | | | | |
| | 8.60 | 4.56 | 4.56 | 7.41 | 6.81 | 68.06 | | | | | | | | | |
| 78 | 104 | 101 | 76 | 70 | 114 | 108 | 1,000 | | | | | | | | |
| | 6.61 | 6.42 | 4.83 | 4.45 | 7.25 | 6.87 | 63.57 | | | | | | | | |
| 79 | 138 | 100 | 93 | 75 | 75 | 111 | 101 | 979 | | | | | | | |
| | 8.25 | 5.98 | 5.56 | 4.49 | 4.49 | 6.64 | 6.04 | 58.55 | | | | | | | |
| 80 | 79 | 115 | 84 | 83 | 66 | 67 | 103 | 93 | 947 | | | | | | |
| | 4.83 | 7.03 | 5.13 | 5.07 | 4.03 | 4.09 | 6.29 | 5.68 | 57.85 | | | | | | |
| 81 | 83 | 77 | 99 | 70 | 73 | 61 | 58 | 99 | 86 | 913 | | | | | |
| | 5.13 | 4.76 | 6.11 | 4.32 | 4.51 | 3.77 | 3.58 | 6.11 | 5.31 | 56.39 | | | | | |
| 82 | 212 | 62 | 51 | 67 | 58 | 55 | 44 | 45 | 76 | 69 | 811 | | | | |
| | 13.68 | 4.00 | 3.29 | 4.32 | 3.74 | 3.55 | 2.84 | 2.90 | 4.90 | 4.45 | 52.32 | | | | |
| 83 | 37 | 75 | 47 | 33 | 49 | 51 | 45 | 35 | 39 | 60 | 83 | 708 | | | |
| | 2.93 | 5.94 | 3.72 | 2.61 | 3.88 | 4.04 | 3.57 | 2.77 | 3.09 | 4.75 | 6.58 | 56.10 | | | |
| 84 | 89 | 91 | 64 | 54 | 32 | 52 | 44 | 42 | 34 | 33 | 62 | 71 | 655 | | |
| | 6.73 | 6.88 | 4.84 | 4.08 | 2.42 | 3.93 | 3.33 | 3.17 | 2.57 | 2.49 | 4.69 | 5.37 | 49.51 | | |
| 85 | 70 | 80 | 79 | 51 | 48 | 30 | 50 | 40 | 41 | 37 | 34 | 59 | 70 | 622 | |
| | 5.34 | 6.10 | 6.02 | 3.89 | 3.66 | 2.29 | 3.81 | 3.05 | 3.13 | 2.82 | 2.59 | 4.50 | 5.34 | 47.41 | |
| 86 | 42 | 57 | 59 | 27 | 45 | 43 | 24 | 47 | 41 | 44 | 36 | 52 | 50 | 57 | 578 |
| | 3.49 | 4.74 | 4.90 | 2.24 | 3.74 | 3.57 | 2.00 | 3.91 | 3.41 | 3.66 | 2.99 | 4.32 | 4.16 | 4.74 | 48.05 |

Table 8, Con't

Industry 38, 1973-1986

| Yr | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 72 | 347 | | | | | | | | | | | | | | |
| | 100.0 | | | | | | | | | | | | | | |
| 73 | 50 | 328 | | | | | | | | | | | | | |
| | 13.23 | 86.77 | | | | | | | | | | | | | |
| 74 | 42 | 40 | 319 | | | | | | | | | | | | |
| | 10.47 | 9.98 | 79.55 | | | | | | | | | | | | |
| 75 | 31 | 27 | 30 | 294 | | | | | | | | | | | |
| | 8.12 | 7.07 | 7.85 | 76.96 | | | | | | | | | | | |
| 76 | 23 | 33 | 26 | 32 | 287 | | | | | | | | | | |
| | 5.74 | 8.23 | 6.48 | 7.98 | 71.57 | | | | | | | | | | |
| 77 | 58 | 26 | 27 | 24 | 34 | 278 | | | | | | | | | |
| | 12.98 | 5.82 | 6.04 | 5.37 | 7.61 | 62.19 | | | | | | | | | |
| 78 | 28 | 32 | 22 | 25 | 26 | 32 | 265 | | | | | | | | |
| | 6.51 | 7.44 | 5.12 | 5.81 | 6.05 | 7.44 | 61.63 | | | | | | | | |
| 79 | 49 | 35 | 27 | 20 | 26 | 26 | 31 | 254 | | | | | | | |
| | 10.47 | 7.48 | 5.77 | 4.27 | 5.56 | 5.56 | 6.62 | 54.27 | | | | | | | |
| 80 | 24 | 41 | 32 | 25 | 20 | 25 | 26 | 28 | 251 | | | | | | |
| | 5.08 | 8.69 | 6.78 | 5.30 | 4.24 | 5.30 | 5.51 | 5.93 | 53.18 | | | | | | |
| 81 | 30 | 24 | 39 | 28 | 21 | 21 | 23 | 21 | 27 | 236 | | | | | |
| | 6.38 | 5.11 | 8.30 | 5.96 | 4.47 | 4.47 | 4.89 | 4.47 | 5.74 | 50.21 | | | | | |
| 82 | 65 | 28 | 20 | 35 | 27 | 16 | 21 | 23 | 19 | 29 | 227 | | | | |
| | 12.75 | 5.49 | 3.92 | 6.86 | 5.29 | 3.14 | 4.12 | 4.51 | 3.73 | 5.69 | 44.51 | | | | |
| 83 | 18 | 33 | 24 | 19 | 34 | 26 | 14 | 22 | 21 | 16 | 33 | 217 | | | |
| | 3.77 | 6.92 | 5.03 | 3.98 | 7.13 | 5.45 | 2.94 | 4.61 | 4.40 | 3.35 | 6.92 | 45.49 | | | |
| 84 | 44 | 33 | 32 | 20 | 17 | 37 | 23 | 12 | 21 | 21 | 14 | 30 | 206 | | |
| | 8.63 | 6.47 | 6.27 | 3.92 | 3.33 | 7.25 | 4.51 | 2.35 | 4.12 | 4.12 | 2.75 | 5.88 | 40.39 | | |
| 85 | 24 | 33 | 31 | 28 | 17 | 15 | 37 | 19 | 13 | 20 | 22 | 11 | 23 | 194 | |
| | 4.93 | 6.78 | 6.37 | 5.75 | 3.49 | 3.08 | 7.60 | 3.90 | 2.67 | 4.11 | 4.52 | 2.26 | 4.72 | 39.84 | |
| 86 | 29 | 25 | 25 | 20 | 26 | 20 | 17 | 34 | 17 | 14 | 21 | 23 | 11 | 22 | 177 |
| | 6.03 | 5.20 | 5.20 | 4.16 | 5.41 | 4.16 | 3.53 | 7.07 | 3.53 | 2.91 | 4.37 | 4.78 | 2.29 | 4.57 | 36.80 |

Table 9
LRD, Industries 35, 36, and 38

| Discontinuous Plant-Year Observations with Coverage Code Values of Zero | | |
|---|---|---|
| Year | All Plants | Dubious Gaps |
| 1974 | 14.53% | 58.06% |
| 1975 | 15.08 | 52.94 |
| 1976 | 14.65 | 71.13 |
| 1977 | 15.77 | 73.76 |
| 1978 | 13.77 | 72.95 |
| 1979 | 13.12 | 69.47 |
| 1980 | 12.97 | 60.34 |
| 1981 | 14.17 | 62.96 |
| 1982 | 16.20 | 45.28 |
| 1983 | 14.80 | 40.30 |
| 1984 | 15.30 | 42.81 |
| 1985 | 15.61 | 66.22 |
| 1986 | 16.70 | 34.30 |

Table 10.1
LRD, Industry 35
Establishments with 250 or More Employees

| Year | 351 | 352 | 353 | 354 | 355 | 356 | 357 | 358 | 359 | All |
|---|---|---|---|---|---|---|---|---|---|---|
| 72 | 65 | 82 | 235 | 172 | 161 | 252 | 150 | 163 | 42 | 1322 |
| 73 | 69 | 94 | 250 | 188 | 169 | 268 | 160 | 183 | 48 | 1429 |
| 74 | 75 | 104 | 266 | 196 | 174 | 294 | 186 | 183 | 49 | 1527 |
| 75 | 74 | 98 | 261 | 183 | 156 | 271 | 174 | 148 | 47 | 1412 |
| 76 | 74 | 94 | 260 | 169 | 156 | 274 | 178 | 155 | 44 | 1404 |
| 77 | 79 | 101 | 277 | 194 | 146 | 293 | 207 | 168 | 47 | 1512 |
| 78 | 78 | 97 | 304 | 199 | 144 | 300 | 225 | 175 | 51 | 1573 |
| 79 | 79 | 115 | 315 | 201 | 146 | 324 | 260 | 174 | 58 | 1672 |
| 80 | 71 | 102 | 308 | 199 | 150 | 328 | 266 | 164 | 49 | 1637 |
| 81 | 74 | 91 | 312 | 187 | 143 | 322 | 274 | 164 | 52 | 1619 |
| 82 | 75 | 75 | 283 | 165 | 130 | 302 | 308 | 157 | 55 | 1550 |
| 83 | 65 | 67 | 184 | 122 | 102 | 242 | 280 | 152 | 48 | 1262 |
| 84 | 68 | 66 | 184 | 126 | 115 | 263 | 287 | 160 | 54 | 1323 |
| 85 | 69 | 64 | 180 | 137 | 116 | 258 | 268 | 165 | 55 | 1312 |
| 86 | 65 | 58 | 149 | 132 | 104 | 243 | 244 | 156 | 52 | 1203 |

Table 10.2
LRD, Industry 36
Establishments with 250 or More Employees

| Year | 361 | 362 | 363 | 364 | 365 | 366 | 367 | 369 | All |
|---|---|---|---|---|---|---|---|---|---|
| 72 | 107 | 184 | 146 | 172 | 72 | 261 | 278 | 101 | 1321 |
| 73 | 119 | 200 | 145 | 174 | 70 | 269 | 303 | 118 | 1398 |
| 74 | 121 | 218 | 147 | 190 | 74 | 273 | 297 | 124 | 1444 |
| 75 | 99 | 187 | 139 | 138 | 65 | 254 | 246 | 101 | 1229 |
| 76 | 97 | 194 | 147 | 157 | 58 | 256 | 270 | 116 | 1295 |
| 77 | 118 | 210 | 144 | 153 | 70 | 290 | 301 | 148 | 1434 |
| 78 | 121 | 217 | 150 | 163 | 68 | 302 | 303 | 151 | 1475 |
| 79 | 122 | 217 | 149 | 174 | 70 | 328 | 348 | 156 | 1564 |
| 80 | 124 | 214 | 147 | 154 | 60 | 327 | 361 | 151 | 1538 |
| 81 | 120 | 212 | 144 | 152 | 52 | 332 | 358 | 153 | 1523 |
| 82 | 114 | 209 | 133 | 143 | 43 | 373 | 416 | 144 | 1575 |
| 83 | 105 | 185 | 132 | 133 | 37 | 360 | 396 | 142 | 1490 |
| 84 | 107 | 197 | 128 | 151 | 38 | 377 | 477 | 156 | 1631 |
| 85 | 101 | 184 | 124 | 153 | 40 | 384 | 447 | 151 | 1584 |
| 86 | 96 | 184 | 122 | 143 | 38 | 378 | 435 | 145 | 1541 |

Table 10.3
LRD, Industry 38
Establishments with 250 or More Employees

| Year | 381 | 382 | 383 | 384 | 385 | 386 | 387 | All |
|---|---|---|---|---|---|---|---|---|
| 72 | 27 | 125 | 14 | 87 | 20 | 43 | 31 | 347 |
| 73 | 38 | 135 | 18 | 90 | 21 | 45 | 31 | 378 |
| 74 | 45 | 138 | 20 | 97 | 24 | 45 | 32 | 401 |
| 75 | 36 | 132 | 23 | 98 | 20 | 43 | 30 | 382 |
| 76 | 37 | 130 | 29 | 101 | 23 | 48 | 33 | 401 |
| 77 | 37 | 164 | 36 | 105 | 26 | 48 | 31 | 447 |
| 78 | 37 | 153 | 32 | 111 | 25 | 46 | 26 | 430 |
| 79 | 37 | 171 | 42 | 113 | 26 | 50 | 29 | 468 |
| 80 | 37 | 175 | 46 | 114 | 27 | 47 | 26 | 472 |
| 81 | 35 | 175 | 42 | 121 | 24 | 50 | 23 | 470 |
| 82 | 39 | 190 | 49 | 137 | 26 | 52 | 17 | 510 |
| 83 | 35 | 174 | 50 | 134 | 19 | 48 | 17 | 477 |
| 84 | 37 | 189 | 47 | 149 | 23 | 47 | 18 | 510 |
| 85 | 37 | 179 | 54 | 146 | 23 | 38 | 10 | 487 |
| 86 | 37 | 171 | 51 | 148 | 24 | 38 | 12 | 481 |

Table 11
Pooled Distributions of Total Employment, 1972-1986
Selected Three-Digit Industry Groups

| | Plants | Percent | Plants | Percent | Plants | Percent |
|---|---|---|---|---|---|---|
| Employees | Industry 353 | | Industry 356 | | Industry 357 | |
| < 250 | - | - | 15 | 0.35 | - | - |
| 250-300 | 546 | 14.49 | 732 | 17.23 | 348 | 10.04 |
| 300-350 | 446 | 11.84 | 609 | 14.33 | 323 | 9.32 |
| 350-450 | 674 | 17.89 | 832 | 19.58 | 468 | 13.50 |
| 450-550 | 489 | 12.98 | 518 | 12.19 | 373 | 10.76 |
| 550-650 | 351 | 9.32 | 359 | 8.45 | 260 | 7.50 |
| 650-750 | 264 | 7.01 | 260 | 6.12 | 179 | 5.16 |
| 750-850 | 189 | 5.02 | 183 | 4.31 | 165 | 4.76 |
| 850-950 | 117 | 3.11 | 146 | 3.44 | 141 | 4.07 |
| 950-1,500 | 303 | 8.04 | 387 | 9.11 | 560 | 16.15 |
| 1,500-3,500 | 294 | 7.80 | 187 | 4.40 | 439 | 12.66 |
| 3,500 + | 95 | 2.52 | 21 | 0.49 | 211 | 6.09 |
| | Industry 362 | | Industry 366 | | Industry 367 | |
| < 250 | - | - | - | - | 14 | 0.30 |
| 250-300 | 447 | 14.84 | 532 | 11.17 | 801 | 15.25 |
| 300-350 | 294 | 9.76 | 401 | 8.42 | 584 | 11.12 |
| 350-450 | 549 | 18.23 | 595 | 12.49 | 916 | 17.44 |
| 450-550 | 404 | 13.41 | 347 | 7.28 | 683 | 13.00 |
| 550-650 | 282 | 9.36 | 321 | 6.74 | 439 | 8.36 |
| 650-750 | 217 | 7.20 | 268 | 5.63 | 340 | 6.47 |
| 750-850 | 195 | 6.47 | 206 | 4.32 | 227 | 4.32 |
| 850-950 | 151 | 5.01 | 193 | 4.05 | 174 | 3.31 |
| 950-1,500 | 308 | 10.23 | 614 | 12.89 | 503 | 9.58 |
| 1,500-3,500 | 138 | 4.58 | 803 | 16.86 | 383 | 7.29 |
| 3,500 + | 27 | 0.90 | 484 | 10.16 | 186 | 3.54 |

Table 12
Pooled Distributions of Total Value of Shipments, 1972-1986
Thousands of 1972 Dollars
Selected Three-Digit Industry Groups

| Value of Shipments | Plants | Percent | Plants | Percent | Plants | Percent |
|---|---|---|---|---|---|---|
| | Industry 353 | | Industry 356 | | Industry 357 | |
| < $10,000 | 652 | 17.30 | 1,131 | 26.62 | 219 | 6.32 |
| $10,000-20,000 | 1,496 | 39.70 | 1,728 | 40.67 | 540 | 15.58 |
| $20,000-30,000 | 596 | 15.82 | 636 | 14.97 | 383 | 11.05 |
| $30,000-40,000 | 279 | 7.40 | 305 | 7.18 | 343 | 9.89 |
| $40,000-50,000 | 182 | 4.83 | 184 | 4.33 | 235 | 6.78 |
| $50,000-75,000 | 229 | 6.08 | 169 | 3.98 | 393 | 11.34 |
| $75,000-125,000 | 197 | 5.23 | 84 | 1.98 | 475 | 13.70 |
| $125,000-175,000 | 53 | 1.41 | * | * | 258 | 7.44 |
| $175,000-225,000 | 20 | 0.53 | 12 | 0.29 | 146 | 4.21 |
| $225,000 and over | 64 | 1.70 | - | - | 475 | 13.70 |
| | Industry 362 | | Industry 366 | | Industry 367 | |
| < $10,000 | 912 | 30.28 | 801 | 16.81 | 1,546 | 29.44 |
| $10,000-20,000 | 1,086 | 36.06 | 1,102 | 23.13 | 1,643 | 31.28 |
| $20,000-30,000 | 512 | 17.00 | 664 | 13.94 | 620 | 11.81 |
| $30,000-40,000 | 223 | 7.40 | 426 | 8.94 | 322 | 6.13 |
| $40,000-50,000 | 83 | 2.76 | 317 | 6.56 | 222 | 4.23 |
| $50,000-75,000 | 89 | 2.95 | 427 | 8.96 | 299 | 5.69 |
| $75,000-125,000 | 58 | 1.93 | 411 | 8.63 | 269 | 5.12 |
| $125,000-175,000 | 29 | 0.96 | 208 | 4.37 | 78 | 1.49 |
| $175,000-225,000 | * | * | 153 | 3.21 | 37 | 0.70 |
| $225,000 and over | 20 | 0.66 | 255 | 5.35 | 216 | 4.11 |

* Indicates Cell Pooled with Preceding Cell
\- Indicates No Observations in the Cell

Table 13
Pooled Distributions of Total Value Added, 1972-1986
Thousands of 1972 Dollars
Selected Three-Digit Industry Groups

| Value Added | Plants | Percent | Plants | Percent | Plants | Percent |
|---|---|---|---|---|---|---|
| | Industry 353 | | Industry 356 | | Industry 357 | |
| <$5,000 | 778 | 20.63 | 823 | 19.37 | 211 | 6.09 |
| $5,000-10,000 | 1,304 | 34.61 | 1,684 | 39.63 | 527 | 15.20 |
| $10,000-20,000 | 904 | 23.99 | 1,160 | 23.30 | 718 | 20.71 |
| $20,000-30,000 | 322 | 8.55 | 319 | 7.51 | 439 | 12.66 |
| $30,000-40,000 | 148 | 3.93 | 125 | 2.94 | 276 | 7.69 |
| $40,000-50,000 | 86 | 2.28 | 54 | 1.27 | 207 | 5.97 |
| $50,000-75,000 | 99 | 2.63 | 69 | 1.62 | 304 | 8.77 |
| $75,000-100,000 | 46 | 1.22 | 15 | 0.35 | 196 | 5.65 |
| $100,000-225,000 | 67 | 1.78 | - | - | 346 | 9.98 |
| $225,000 and over | 14 | 0.37 | - | - | 243 | 7.01 |
| | Industry 362 | | Industry 366 | | Industry 367 | |
| <$5,000 | 766 | 25.43 | 573 | 12.03 | 1,093 | 20.81 |
| $5,000-10,000 | 1,045 | 34.69 | 1,062 | 22.29 | 1,586 | 30.20 |
| $10,000-20,000 | 801 | 25.59 | 1,203 | 21.47 | 1,186 | 22.58 |
| $20,000-30,000 | 204 | 6.77 | 526 | 11.04 | 434 | 8.26 |
| $30,000-40,000 | 78 | 2.59 | 358 | 7.51 | 253 | 4.83 |
| $40,000-50,000 | 28 | 0.93 | 200 | 4.70 | 149 | 2.84 |
| $50,000-75,000 | 48 | 1.59 | 370 | 7.77 | 180 | 3.43 |
| $75,000-100,000 | 21 | 0.70 | 198 | 4.16 | 84 | 1.60 |
| $100,000-225,000 | 21 | 0.70 | 372 | 7.81 | 137 | 2.61 |
| $225,000 and over | - | - | 82 | 1.72 | 150 | 2.86 |

- Indicates No Observations in the Cell

Table 14
Pooled Distributions of Total Capital Stock, 1972-1986
Thousands of 1972 Dollars
Selected Three-Digit Industry Groups

| Capital Stock | Plants | Percent | Plants | Percent | Plants | Percent |
|---|---|---|---|---|---|---|
| | Industry 353 | | Industry 356 | | Industry 357 | |
| < $2,500 | 112 | 2.97 | 84 | 1.98 | 139 | 4.01 |
| $2,500-5,000 | 229 | 6.08 | 98 | 2.31 | 317 | 9.14 |
| $5,000-10,000 | 821 | 21.79 | 421 | 9.94 | 547 | 15.78 |
| $10,000-20,000 | 1,016 | 29.96 | 1,016 | 24.00 | 767 | 22.12 |
| $20,000-35,000 | 764 | 20.28 | 957 | 22.60 | 587 | 16.93 |
| $35,000-50,000 | 353 | 9.37 | 550 | 12.99 | 245 | 7.07 |
| $50,000-75,000 | 153 | 4.06 | 537 | 12.68 | 262 | 7.56 |
| $75,000-100,000 | 107 | 2.84 | 232 | 5.48 | 141 | 4.07 |
| $100,000-150,000 | 117 | 3.11 | 186 | 4.39 | 169 | 4.87 |
| $150,000-250,000 | 72 | 1.91 | 112 | 2.65 | 133 | 3.84 |
| $250,000 and over | 24 | 0.64 | 41 | 0.97 | 160 | 4.61 |
| | Industry 362 | | Industry 366 | | Industry 367 | |
| < $2,500 | 36 | 1.20 | 345 | 7.25 | 90 | 1.72 |
| $2,500-5,000 | 179 | 5.94 | 970 | 20.38 | 181 | 3.46 |
| $5,000-10,000 | 785 | 26.06 | 1,040 | 21.85 | 704 | 13.45 |
| $10,000-20,000 | 732 | 24.30 | 973 | 20.45 | 1,115 | 21.29 |
| $20,000-35,000 | 574 | 19.06 | 520 | 10.93 | 681 | 13.01 |
| $35,000-50,000 | 208 | 6.91 | 311 | 6.53 | 510 | 9.74 |
| $50,000-75,000 | 188 | 6.24 | 275 | 5.78 | 622 | 11.88 |
| $75,000-100,000 | 91 | 3.02 | 101 | 2.12 | 369 | 7.05 |
| $100,000-150,000 | 105 | 3.49 | 135 | 2.84 | 382 | 7.30 |
| $150,000-250,000 | 95 | 3.15 | 78 | 1.64 | 252 | 4.81 |
| $250,000 and over | 19 | 0.63 | 11 | 0.23 | 330 | 6.30 |

Example from Cochran on increased efficiency in estimates of change and percent change with repeated sampling:

| Estimates for $\bar{y} = f\bar{y}_m + (1-f)\bar{y}_u$, where $m, u$ denote matched and unmatched observations respectively | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $r = 0.7$ | | $r = 0.8$ | | $r = 0.9$ | | $r = 0.95$ | |
| $h$ | $\frac{m}{n} = \frac{1}{2}$, $f = 0.35$ | $\frac{m}{n} = \frac{3}{4}$, $f = 0.2$ | $\frac{m}{n} = \frac{1}{2}$, $f = 0.35$ | $\frac{m}{n} = \frac{3}{4}$, $f = 0.2$ | $\frac{m}{n} = \frac{1}{2}$, $f = 0.35$ | $\frac{m}{n} = \frac{3}{4}$, $f = 0.2$ | $\frac{m}{n} = \frac{1}{2}$, $f = 0.35$ | $\frac{m}{n} = \frac{3}{4}$, $f = 0.2$ |
| Percent Gain in Efficiency for the Current Estimate | | | | | | | | |
| 2 | 14 | 10 | 22 | 14 | 33 | 19 | 41 | 22 |
| 3 | 16 | 14 | 30 | 20 | 52 | 32 | 67 | 39 |
| 4 | 17 | 15 | 32 | 24 | 59 | 40 | 79 | 52 |
| Percent Gain in Efficiency for the Estimate of Change | | | | | | | | |
| 2 | 106 | 153 | 156 | 233 | 245 | 399 | 326 | 565 |
| 3 | 113 | 160 | 170 | 245 | 277 | 415 | 365 | 588 |
| 4 | 115 | 160 | 174 | 251 | 285 | 424 | 388 | 603 |
| Source: Cochran, Sampling Techniques (1977), p. 353, Table 12.5 | | | | | | | | |